

T9519

STIC-Biotech/ChemLib

From: Slobodyansky, Elizabeth
Sent: Monday, September 13, 1999 5:22 PM
To: STIC-Biotech/ChemLib
Subject: 09/068,507

This is a date case, please let me know if I need to do anything to facilitate.

Please search for case **09/068,507**:

1. Seq ID NOs 1 + 2 against commercial ~~and interference~~ databases.

Thank you very much.

Elizabeth Slobodyansky

AM 1652
CM1-10D11
tel. 306-3222

P

IMPORTANT INFORMATION ABOUT YOUR SEQUENCE SEARCH:

Compugen Sequence searching hardware and software explained:

This is the new sequence searching system that is currently being phased into as a replacement for the Maspar/Mpsrch platform. This system has been tested by both searchers and examiners, and has shown equivalent results to the Maspar system for the same databases. The results output format for all Compugen printed results are essentially the same except for translations.

Translation searching on Compugen explained:

The Compugen system utilizes Framesearch software for translations of proteins to nucleotides, and nucleotides to proteins. Some examiners have found these to be superior to the backtranslate software on Maspars.

FrameSearch searches a group of protein sequences for similarity to one or more nucleotide query sequences, or searches a group of nucleotide sequences for similarity to one or more protein query sequences. For each sequence comparison, the program finds an optimal alignment between the protein sequence and the corresponding codons on each the nucleotide sequence. Optimal alignments may include reading frame shifts. Please see any of the professional searching staff if you need assistance with this format.

File extensions for Compugen results transferred to floppy disks.

Compugen system search results will be delivered in one of two possible formats:

1. Standard concatenated files with .flp extension.
2. Compressed .zip files which decompressed yield two files as described below:

US08123456.cmr - Contains all commercial databases, may include Issued
US08123456.pen - Contains pending file results only

VERY IMPORTANT NOTE ABOUT PENDING FILE SEARCHES.

If your search contains file names with the following bolded extensions:

US08123456.rap US08123456.rnp

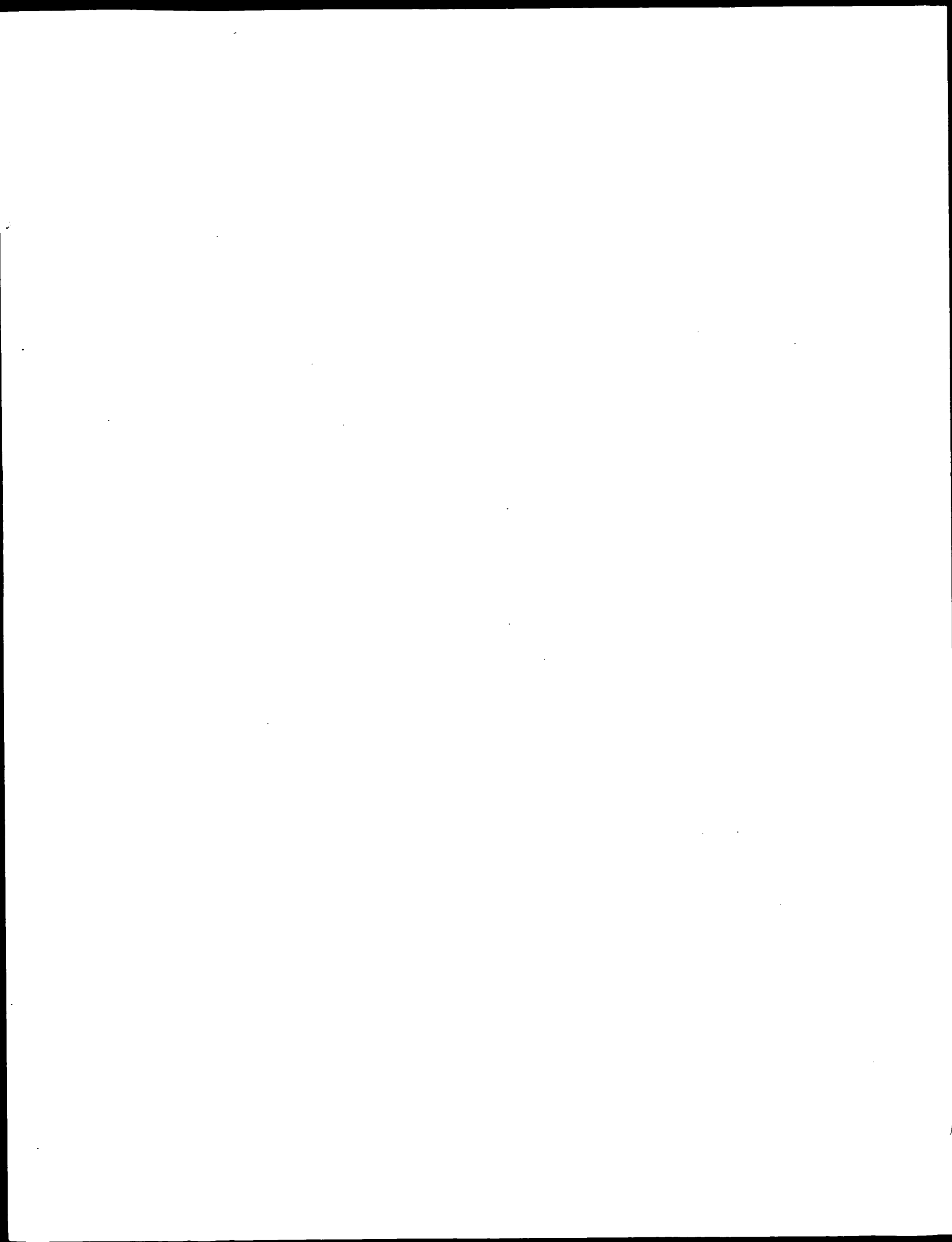
Do not leave this search in the case, during prosecution, or after the case issues, since it contains pending data which is confidential.

QUESTIONS? Contact any of the following:

Dilip Pandya, Chief, Information Branch, 308-4268

Professional searching staff:

John Dantzman (308-4488); Jan Delaval (308-4498); Mary Hale (308-4258); Barb O'Bryen (308-4291); David Schreiber (308-4292); Paula Sheppard (308-4499); Mark Spencer (308-4266); Beverly Shears (308-4994); Alex Wacławiw (308-4491).



FOR OFFICIAL USE ONLY

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office195-19
SEARCH REQUEST FORM

Examiner # (Mandatory): _____ Requester's Full Name: _____

Art Unit _____ Location (Bldg/Room#): 10011 Phone (circle 305 306 308) _____

Serial Number: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Keywords (include any known synonyms registry numbers, explanation of initialisms): _____

Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

FOR OFFICIAL USE ONLY

STAFF USE ONLY

Searcher: BOBSearcher Phone #: 303-4740

Searcher Location: _____

Date Picked Up: 9-15

Date Completed: _____

Clerical Prep Time: _____

Terminal Time: _____

Number of Databases: _____

Type of Search

1 N.A. Sequence1 A.A. Sequence

____ Structure (#)

____ Bibliographic

____ Litigation1

____ Fulltext

____ Procurement

____ Other

Vendors (include cost where applicable)

____ STN

____ Questel/Orbit

____ Lexis/Nexis

____ WWW/Internet

AP 2 In-house sequence systems (list)

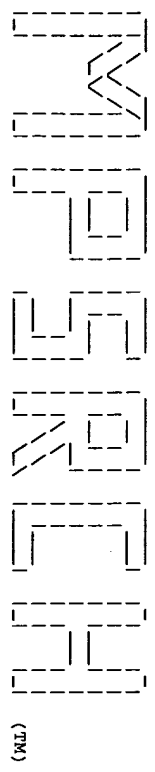
____ Dialog

____ Dr. Link

____ Westlaw

____ Other (specify)





Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSPRCH_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 15 10:37:50 1999; Maspar time 11.18 Seconds
Tabular output not generated. 49.467 Million cell updates/sec

Title: >US-09-068-507A-1
Description: (1-26) from US09068507A.pep
Perfect Score: 199
Sequence: 1 KSSAYSLQMGATAIKQYKLFKKWGW 26

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 22.683; Variance 77.937; scale 0.291

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	199	100.0	26	24	W17989	Bacteriocin-inducing amino acid sequence	2.87e-13
2	77	38.7	832	33	W69847	Human GABA-BR1a/b rec	4.40e+00
3	75	37.7	793	29	W40117	Rat GABA-BR1b recepto	6.84e+00
4	75	37.7	844	29	W40118	Human GABA-BR1b recepto	6.84e+00
5	75	37.7	844	29	W40119	Rat GABA-BR1a recepto	6.84e+00
6	75	37.7	960	29	W40116	Human GABA-BR1a recepto	6.84e+00
7	71	35.7	572	13	R78520	Partial ALK protein.	1.64e+01
8	67	33.7	234	34	W62742	Streptococcus pneumoniae	3.87e+01
9	64	32.2	115	31	W28166	Staphylococcus aureus	7.28e+01
10	62	31.2	467	28	W41077	T. litoralis ADP/AT	1.10e+02
11	62	31.2	572	8	R39284	Parainfluenza virus t	1.10e+02
12	62	31.2	1577	16	R91047	Alpha-D-glucosyltrans	1.10e+02
13	60	30.2	13	24	W12897	Antimicrobial cationi	1.67e+02
14	60	30.2	74	18	R95262	Bacteriocin derivativ	1.67e+02
15	60	30.2	128	33	W61021	Streptococcus pneumoniae	1.67e+02
16	60	30.2	305	26	W23077	Sulfolobus solfataric	1.67e+02

17	60	30.2	558	3	P50501	Sequence of bovine pa	1.67e+02
18	60	30.2	572	15	R79037	PI-3 haemagglutinin	1.67e+02
19	60	30.2	572	1	P94799	Perdue strain of tran	1.67e+02
20	60	30.2	572	1	R06023	Viral haemagglutinin	1.67e+02
21	60	30.2	572	1	P94800	Parainfluenzae-3 gene	1.67e+02
22	60	30.2	572	6	R31957	Sequence encoded by p	1.67e+02
23	59	29.6	363	4	R22551	T lymphocyte-specific	2.04e+02
24	59	29.6	370	18	R98127	Human lymphocyte cell	2.04e+02
25	59	29.6	370	18	R98128	Human lymphocyte cell	2.04e+02
26	59	29.6	371	18	R98124	Human lymphocyte cell	2.04e+02
27	59	29.6	371	18	R98114	Human lymphocyte cell	2.04e+02
28	59	29.6	371	18	R98113	Human lymphocyte cell	2.04e+02
29	59	29.6	371	18	R98112	Human lymphocyte cell	2.04e+02
30	59	29.6	371	18	R98129	Human lymphocyte cell	2.04e+02
31	59	29.6	371	18	R98118	Human lymphocyte cell	2.04e+02
32	59	29.6	371	18	R98121	Human lymphocyte cell	2.04e+02
33	59	29.6	372	31	W37781	Homo sapiens lymphoc	2.04e+02
34	59	29.6	372	8	R38908	HuDR.	2.04e+02
35	59	29.6	372	4	R24026	Sequence of human lym	2.04e+02
36	59	29.6	372	37	W73264	Human lymphocyte homi	2.04e+02
37	59	29.6	374	18	R98134	Human lymphocyte cell	2.04e+02
38	59	29.6	374	18	R98132	Human lymphocyte cell	2.04e+02
39	59	29.6	384	1	R06645	Lymphocyte associated	2.04e+02
40	59	29.6	385	11	R56653	L-selectin.	2.04e+02
41	59	29.6	385	7	R34197	Sequence encoded by L	2.04e+02
42	59	29.6	385	18	R91442	Human Leu8 antigen.	2.04e+02
43	59	29.6	797	30	W36508	Human RENT1 protein.	2.04e+02
44	59	29.6	1043	30	W36509	Murine RENT1 protein.	2.04e+02
45	59	29.6	1462	7	R37508	Human DNA polymerase	2.04e+02

ALIGNMENTS

RESULT 1
ID W17989 standard; Peptide; 26 AA.
AC W17989;
DT 10-DEC-1997 (first entry)
DE Bacteriocin-inducing peptide.
KW Gene expression; promoter; Lactic acid bacterium; bacteriocin;
KW vaccine; vector.
OS Lactobacillus plantarum strain C11.
PN W05718316-A1.
PD 22-MAY-1997.
PE 13-NOV-1996; NC00266.
PR 13-NOV-1995; NC004575.
PA (BRUR/) BRURBERG M B.
PA (ELUS/) ELUSINK V G H.
PA (NESI/) NES I F.
PI Brurberg MB, Elusink VGH, Nes IF;
DR WPI: 97-289292/26.
PT Gene expression system providing regulated lactic acid bacteria
PT protein production - uses new inducing peptide involved in
PT bacteriocin synthesis, useful in fermentation and as a drug delivery
PT system
PS Claim 5; Page 28; 39pp; English.
CC This 26-residue peptide from Lactobacillus plantarum strain C11
CC can be used to induce and/or maintain the expression of genes
CC involved in the production of bacteriocins. The invention relates
CC to the discovery of a new regulatory mechanism for gene expression
CC in lactic acid bacteria that includes previously unrecognized,
CC strongly regulatable promoter elements (see T67142-48). Expression
CC of other genes that are preceded by such promoter elements and
CC which are introduced into C11 cells can be induced by adding the
CC (chemically synthesized) 26-residue inducing peptide. A recombinant
CC vector containing such gene expression system, and host cells
CC transformed with this vector, incorporated in the chromosome,
CC and/or having integrated into its chromosome a promoter linked to
CC an integrated gene encoding a specific protein are claimed.
CC Typical applications are in fermentations (e.g. where the gene for
CC an enzyme is regulated) and production of specific proteins, or
CC where the bacteria express a surface antigen, as vaccines.
SQ Sequence 26 AA;

Query Match 100.0%; Score 199; DB 24; Length 26;

Best Local Similarity 100.0%; Pred. No. 2, 87e-13;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 kssayslmgataikvkkfkxgw 26

OY 1 KSSAYSLMGATAIKOVKKLFKKMGW 26

RESULT 2

ID W69847 standard; Protein: 822 AA.

AC W69847;

DT 07-OCT-1998 (first entry)

DE Amino acid sequence of p96, a protein present in SA-17S complex.

KM p96 protein: secretion associated 17S complex; SA-17S.

KM syntaxin-containing complex; SC complex; screening; modulate;

KM vesicular release; synaptic transmission; secretory process;

KM treatment; affective disorder; depression; manic-depressive disorder;

KM anxiety disorder; neurodegenerative disease; schizophrenia; anaesthesia;

KM hormonal imbalance; antigen processing; ss.

OS Rattus sp.

FT Key Location/Qualifiers

FT Misc.difference 412 /note="not specified"

PN W09828419-A2.

PD 02-JUL-1998.

PF 19-DEC-1997; U23498.

PR 20-DEC-1996; US-033905.

PA (STRD) UNIV LELAND STANFORD JUNIOR.

PI Scheller RH;

DR WPI: 98-377650/32.

DR N-PSDB: V42652.

PT New isolated vesicle secretion associated poly:peptide(s) - used to

PT develop products for treating e.g. effective disorder,

PT neurodegenerative disease, hormone imbalances, immune system

PT disorders or tumours

PS Claim 1d: Pages 107-109; 133pp; English.

CC The present sequence represents a p96 protein. This protein is present

CC the secretion associated 17S (SA-17S) complex. Eight proteins form the

CC SA-17S complex, which binds a syntaxin-containing (SC) complex. The

CC SA-17S polypeptides and nucleotide sequences encoding them can be used

CC for screening for compounds which modulate vesicular release involved in

CC synaptic transmission and other secretory processes. Compounds which

CC enhance binding between the SA-17S and SC complexes may be used to treat

CC an affective disorder such as depression, manic-depressive disorders and

CC anxiety disorders, or a neurodegenerative disease such as Parkinson's

CC disease or Huntington's disease. Compounds which inhibit binding between

CC the SA-17S and SC complexes may be used to treat a disorder of thought,

CC such as schizophrenia, or for anaesthesia. The compounds can also be used,

CC to intervene in the endocrine system for treatment of hormonal

CC imbalances, the immune system for intervention in antigen processing,

CC secreted immunomodulators, and viral processing, as well as

CC anti-tumour applications, such as regulation of membrane trafficking

CC during rapid cell division.

CC Sequence 822 AA;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

KM epilepsy; cognitive function.

OS Homo sapiens.

PN W09746675-A1.

PD 11-DEC-1997.

PF 19-MAR-1997; E01370.

PR 22-NOV-1996; US-756091.

PR 30-MAY-1996; US-655716.

PA (NOVS) NOVARTIS AG.

PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;

DR WPI: 98-042183/04.

DR N-PSDB: V10265.

PT Purified GABA-B receptor or receptor protein - and antagonists of

PT these which may be useful in treating nervous system disorders

PS Claim 4; Page 62-67; 108pp; English.

CC This sequence represents a novel human GABA-B receptor protein,

CC GABA-BR1a/b. GABA (gamma-aminobutyric acid) is the major inhibitory

CC neurotransmitter found in the brain and peripheral nervous system

CC and this receptor may be used for the identification of GABA-B

CC receptor agonists and antagonists. Such proteins may be used in

CC treatment of dementia, depression, anxiety, epilepsy, spasticity,

CC bronchial inflammation or asthma or to improve cognitive function.

CC GABA-B receptor ligands and probes derived from this sequence can be

CC used to assay for GABA-B receptors or DNA encoding them.

CC Sequence 793 AA;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

KM epilepsy; cognitive function.

OS Homo sapiens.

PN W09746675-A1.

PD 11-DEC-1997.

PF 19-MAR-1997; E01370.

PR 22-NOV-1996; US-756091.

PR 30-MAY-1996; US-655716.

PA (NOVS) NOVARTIS AG.

PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;

DR WPI: 98-042183/04.

DR N-PSDB: V10265.

PT Purified GABA-B receptor or receptor protein - and antagonists of

PT these which may be useful in treating nervous system disorders

PS Claim 4; Page 62-67; 108pp; English.

CC This sequence represents a novel human GABA-B receptor protein,

CC GABA-BR1a/b. GABA (gamma-aminobutyric acid) is the major inhibitory

CC neurotransmitter found in the brain and peripheral nervous system

CC and this receptor may be used for the identification of GABA-B

CC receptor agonists and antagonists. Such proteins may be used in

CC treatment of dementia, depression, anxiety, epilepsy, spasticity,

CC bronchial inflammation or asthma or to improve cognitive function.

CC GABA-B receptor ligands and probes derived from this sequence can be

CC used to assay for GABA-B receptors or DNA encoding them.

CC Sequence 793 AA;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

KM epilepsy; cognitive function.

OS Homo sapiens.

PN W09746675-A1.

PD 11-DEC-1997.

PF 19-MAR-1997; E01370.

PR 22-NOV-1996; US-756091.

PR 30-MAY-1996; US-655716.

PA (NOVS) NOVARTIS AG.

PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;

DR WPI: 98-042183/04.

DR N-PSDB: V10265.

PT Purified GABA-B receptor or receptor protein - and antagonists of

PT these which may be useful in treating nervous system disorders

PS Claim 4; Page 62-67; 108pp; English.

CC This sequence represents a novel human GABA-B receptor protein,

CC GABA-BR1a/b. GABA (gamma-aminobutyric acid) is the major inhibitory

CC neurotransmitter found in the brain and peripheral nervous system

CC and this receptor may be used for the identification of GABA-B

CC receptor agonists and antagonists. Such proteins may be used in

CC treatment of dementia, depression, anxiety, epilepsy, spasticity,

CC bronchial inflammation or asthma or to improve cognitive function.

CC GABA-B receptor ligands and probes derived from this sequence can be

CC used to assay for GABA-B receptors or DNA encoding them.

CC Sequence 793 AA;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

KM epilepsy; cognitive function.

OS Homo sapiens.

PN W09746675-A1.

PD 11-DEC-1997.

PF 19-MAR-1997; E01370.

PR 22-NOV-1996; US-756091.

PR 30-MAY-1996; US-655716.

PA (NOVS) NOVARTIS AG.

PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;

DR WPI: 98-042183/04.

OY 19 KLFKKMGW 26

RESULT 5
ID W40119 standard; Protein: 844 AA.

AC W40119;
DT 03-JUN-1998 (first entry)
DE Human GABA-BR1b receptor protein.
KW Gamma-aminobutyric acid; GABA-BR1a/b receptor; human; brain; agonist;
KW Inhibitory neurotransmitter; peripheral nervous system; antagonist;
KW treatment; dementia; depression; anxiety; bronchial inflammation; asthma;
KW epilepsy; cognitive function.
OS Homo sapiens.
PN W09746675-A1.
PD 11-DEC-1997.
PE 19-MAR-1997; E01370.
PR 22-NOV-1996; US-756091.
PR 30-MAY-1996; US-655716.
PA (NOVS) NOVARTIS AG.
PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;
WPI: 98-042183/04.
DR N-PSDB: V10267.
DR Purified GABA-B receptor or receptor protein - and antagonists of
PT these which may be useful in treating nervous system disorders
PS Claim 4; Page 86-90; 108pp; English.
CC This sequence represents a novel human GABA-B receptor protein,
CC GABA-BR1b. GABA (gamma-aminobutyric acid) is the major inhibitory
CC neurotransmitter found in the brain and peripheral nervous system
CC and this receptor may be used for the identification of GABA-B
CC receptor agonists and antagonists. Such proteins may be used in
CC treatment of dementia, depression, anxiety, epilepsy, spasticity,
CC bronchial inflammation or asthma or to improve cognitive function.
CC GABA-B receptor ligands and probes derived from this sequence can be
CC used to assay for GABA-B receptors or DNA encoding them.
SQ Sequence 844 AA;

Query Match

37.7%; Score 75; DB 29; Length 844;

Best Local Similarity 87.5%; Pred. No. 6,84e+00;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 182 Klfekgw 189

OY 19 KLFKKMGW 26

RESULT 6
ID W40116 standard; Protein: 960 AA.

AC W40116;
DT 03-JUN-1998 (first entry)
DE Rat GABA-BR1a receptor protein.
KW Gamma-aminobutyric acid; GABA-BR1a receptor; rat; brain; agonist;
KW Inhibitory neurotransmitter; peripheral nervous system; antagonist;
KW treatment; dementia; depression; anxiety; bronchial inflammation; asthma;
KW epilepsy; cognitive function.
OS Rattus norvegicus.
PN W09746675-A1.
PD 11-DEC-1997.
PE 19-MAR-1997; E01370.
PR 22-NOV-1996; US-756091.
PR 30-MAY-1996; US-655716.
PA (NOVS) NOVARTIS AG.
PI Bettler B, Bittiger H, Froestl W, Kaupmann K, Mickel SJ;
WPI: 98-042183/04.
DR N-PSDB: V10264.
DR Purified GABA-B receptor or receptor protein - and antagonists of
PT these which may be useful in treating nervous system disorders
PS Claim 4; Page 50-56; 108pp; English.
CC This sequence represents a novel rat GABA-B receptor protein,
CC GABA-BR1a. GABA (gamma-aminobutyric acid) is the major inhibitory
CC neurotransmitter found in the brain and peripheral nervous system
CC and this receptor may be used for the identification of GABA-B
CC receptor agonists and antagonists. Such proteins may be used in
CC treatment of dementia, depression, anxiety, epilepsy, spasticity,

CC bronchial inflammation or asthma or to improve cognitive function.
CC GABA-B receptor ligands and probes derived from this sequence can be
CC used to assay for GABA-B receptors or DNA encoding them.
SQ Sequence 960 AA;

Query Match

37.7%; Score 75; DB 29; Length 960;

Best Local Similarity 87.5%; Pred. No. 6,84e+00;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 298 Klfekgw 305

OY 19 KLFKKMGW 26

RESULT 7
ID R78520 standard; Protein: 572 AA.

AC R78520;
DT 15-NOV-1995 (first entry)
DE Partial ALK protein.
KW ALK protein; protein-tyrosine-kinase; fusion protein; NPM protein;
KW nucleolar phosphoprotein; anaplastic large cell lymphoma;
KW t(2;5) translocation; diagnosis; gene therapy.
OS Homo sapiens.
PN W09515331-A.
PD 08-JUN-1995.
PE 05-DEC-1994; U13947.
PR 03-DEC-1993; US-160861.
PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PI Look AT, Morris SW;
WPI: 95-215226/28.
DR N-PSDB: Q95341.
DR Methods for detecting human t(2;5) lymphoma - for detection and
PT diagnosis of anaplastic large cell lymphoma(s)
PS Claim 21; Page 40-42; 70pp; English.
CC The translocation event that occurs in human t(2;5) lymphoma
CC brings sequences from the nucleolar phosphoprotein (NPM) gene on
CC chromosome 3q35 to those from a protein-tyrosine-kinase (ALK) gene
CC on chromosome-2q23. Nucleic acids encoding a partial ALK protein
CC (R78520) and the ALK/NPM fusion protein (R78521) were isolated
CC (Q95541, Q95542). Identification of the NPM/ALK fusion allows
CC lymphoma diagnosis, or therapy through the use of antisense RNA or
CC ribozymes.
SQ Sequence 572 AA;

Query Match

35.7%; Score 71; DB 13; Length 572;

Best Local Similarity 52.4%; Pred. No. 1,64e+01;

Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 2 slqegatgghscpamkkgw 22

OY 6 SLQMGTAIKQYKDLKKMGW 26

RESULT 8
ID W62742 standard; Protein: 234 AA.

AC W62742;
DT 09-NOV-1998 (first entry)
DE Streptococcus pneumoniae polypeptide.
KW Polypeptide; ORF; open reading frame; infection; bacterial;
KW streptococcal; bacteremia; diagnosis; prophylaxis.
OS Streptococcus pneumoniae.
PN W09823651-A1.
PD 04-JUN-1998.
PE 24-NOV-1997; U21976.
PR 27-NOV-1996; US-031879.
PA (SMK) SMITHKLINE BEECHAM CORP.
PI (SMK) SMITHKLINE BEECHAM PLC.
PI Black MT, Hodgson JE, Knowles DJC, Ionetto MA, Nicholas RO,
WPI: 98-322654/28.
DR Streptococcus pneumoniae polynucleotides - useful for developing
PT products for diagnosis, prevention and treatment of infections e.g.
PT pneumonia, bacteremia, meningitis or endocarditis

RESULT	12
ID	R91047 standard; Protein; 1577 AA.
AC	R91047:
DT	22-MAY-1996 (first entry)
ED	Alpha-D-glucosyltransferase.
KD	Alpha-D-glucosyltransferase.
MD	sucrose; transgenic plant; cloning; Escherichia coli;
PK	phage lambda-cl3; vector; plasmid pSGS502;
SV	gene transfer; crop improvement; storage carbohydrate; pasture;

KM feedstuff; senescence; dextran; binder; food; pharmaceutical.
 OS Streptococcus salivarius strain ATCC 25975.
 PD WO9606173-A1.
 PD 29-FEB-1996.
 PF 24-AUG-1995: AU0527.
 PA (GIRFF/) GIFFARD P M.
 PA (JACOQ/) JACQUES N A.
 PA (SIMP/) SIMPSON C L.
 PI Giffard PM, Jacques NA, Simpson CL;
 DR MPI: 96-151376/15.
 DR N-PSDB: T113139.
 PT Plants contg. new bacterial DNA encoding glucosyl transferase
 PT activity - retain higher levels of stored carbohydrate(s) in a form
 PT readily digestible by ruminants
 PS Claim 4: Page 16-20: 31pp; English.
 CC The sequence represents an alpha-D-glucosyltransferase from
 CC Streptococcus salivarius. The enzyme is primer-independent, and
 CC produces soluble glucan from sucrose. A gene encoding the enzyme
 CC may be cloned and expressed in Escherichia coli using a subclone
 CC of phage lambda-cl3, e.g. plasmid PGSG501 or plasmid PGSG502. The
 CC DNA may also be expressed in a transgenic plant, to improve the
 CC level of stored carbohydrate in a pasture plant which normally
 CC contains low levels, or to prevent degradation of stored carbohydrate
 CC during plant senescence. Dextran may be isolated from the plant, for
 CC use as a food binder or pharmaceutical additive. Primer independence
 CC ensures that the enzyme will be functional in plants. The glucan is
 CC poorly degraded in plants but easily degraded by bacteria in the rumen
 CC of grazing livestock.
 SO Sequence 1577 AA;

Query Match 31.2%; Score 62; DB 16; Length 1577;
 Best Local Similarity 66.7%; Pred. No. 1.10e+02;
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Db 959 laqnaHlfkewg 970
 14 INQVKRLFRKMG 25

RESULT 13
 ID W12897 standard; peptide: 13 AA.
 AC W12897;
 DT 10-DEC-1997 (first entry)
 DE Antimicrobial cationic peptide CP-23.
 KM Bacterial; viral; antitumour; food; preservative; inhibitor; growth;
 KM bacterium; yeast; endotoxaemia; sepsis; antibiotic; fungal;
 KM antiviral; Candida albicans; sterility; Salmonella; Yersinia;
 KM Shigella.
 OS Synthetic.
 PN WO9708199-A2.
 PD 06-MAR-1997.
 PF 23-AUG-1996: IB0996.
 PR 23-AUG-1995: US-002687.
 PA (UTBR-) UNIV BRITISH COLUMBIA.
 PI Falla JV, Gough M, Hancock RM;
 DR MPI: 97-179179/16.
 PT Cationic peptide(s) having anti-microbial activity - used for the
 PT inhibition of bacterial and viral growth, as an antitumour agent,
 PT and as a food preservative
 PS Claim 3: Page 66: 89pp; English.
 CC The present sequence represents a specifically claimed novel isolated
 CC cationic peptide which has antimicrobial activity. The amino acid
 CC sequence of antimicrobial cationic peptides (including the present
 CC sequence) is selected from: X1X1ProX2X3X2Pro(X2X2Pro)X2X3(X5)O;
 CC X1X1ProX2X3X4(X5)ProX2X3X3; X1X1X3(ProTIP)X3X2X5X2X5X2(X5)O;
 CC X1X1X3X3X2Pro(X2X2Pro)X2(X5)m; where m = 1-5; n = 1-2; o = 2-5; r
 CC = 0-8; u = 0-1; x1 = Ile, Leu, Val, Phe, Tyr, Trp or Met; x2 = Trp or
 CC Phe; x3 = Arg or Lys; x4 = Trp or Lys; and x5 = Phe, Trp, Arg, Lys or
 CC Pro. The peptides are preferably amidated or carboxymethylated. The
 CC peptides may be used in methods for inhibiting the growth of a bacterium
 CC or yeast, or for inhibiting an endotoxaemia or sepsis associated
 CC disorder in a subject. The peptides have a broad activity against

CC antibiotic resistant bacteria, combined with activity against the
 CC medically important fungus Candida albicans. In addition, the peptides
 CC are useful as antitumour agents and/or antiviral agents. The peptides
 CC may be used as sterilants or preservatives of materials susceptible to
 CC microbial or viral contamination, e.g. in processed foods to inhibit
 CC Salmonella, Yersinia and Shigella. The peptides are compact and tend to
 CC have a unique polypyrrolone type II extended helix structure that permits
 CC them to span the membrane with relatively few amino acids. The peptides
 CC possess the ability to work synergistically with antibiotics, and in
 CC addition, some of them possess anti-endotoxin activity.
 SO Sequence 13 AA;

Query Match 30.2%; Score 60; DB 24; Length 13;
 Best Local Similarity 71.4%; Pred. No. 1.67e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 1 ffrkpw 7
 20 LFRKMGW 26

RESULT 14
 ID R95262 standard; Protein; 74 AA.
 AC R95262;
 DT 11-NOV-1996 (first entry)
 DE Bacteriocin derivative (lactococcin G beta) peptide C.
 KM Protease; restriction protease; cleavage; proteolytic domain;
 KM protein.
 OS Lactococcus lactis.
 FH Key Location/Qualifiers
 FT region 1..33
 FT /label= N-terminal affinity tag.
 FT cleavage_site 31..32
 FT /label= Enterokinase cleavage site.
 FT peptide 34..74
 FT /label= Mature peptide.
 PN WO9616167-A1.
 PD 30-MAY-1996.
 PF 16-NOV-1995: N00213.
 PR 17-NOV-1994: NO-004411.
 PA (HAYNA) HAVARSTEIN L S.
 PA (NESI/) NES I F.
 PI Havarstein LS, Nes IF;
 DR MPI: 96-268604/27.
 PT Protease(s) with highly specific cleavage activity - have
 PT proteolytic domains derived from N-terminals of ABC transporters,
 PT pref. 150 N-terminal amino acids of Lagd
 PS Example 1; Figure 3C; 43pp; English.
 CC Proteases derived from ABC-transporters containing N-terminal
 CC proteolytic domains are restriction proteases that can be used for
 CC the cleavage of proteins at specific amino acid cleavage sites. The
 CC protease was produced by isolating a 5 kb SpeI DNA fragment
 CC containing the protease N-terminal coding sequence, cloning
 CC this fragment into a pBluescript II SK+ plasmid and then using two
 CC primers (T29648, T29649) to amplify the N-terminal proteolytic
 CC domain. The bacteriocin derivative was produced recombinantly as a
 CC test substrate for the ABC-protease.
 SO Sequence 74 AA;

Query Match 30.2%; Score 60; DB 18; Length 74;
 Best Local Similarity 47.4%; Pred. No. 1.67e+02;
 Matches 9; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 26 qmgardlydddkbpkxgw 44
 8 QMGATAIKQVKLFKRWG 26

RESULT 15
 ID W61021 standard; Protein; 128 AA.
 AC W61021;
 DT 13-OCT-1998 (first entry)
 DE Streptococcus pneumoniae encoded polypeptide.

KM coding region: ORF: open reading frame: antibacterial;
 KM infection: prevention: meningitis.
 OS Streptococcus pneumoniae.
 PN W09819689-A1.
 PD 14-MAY-1998.
 PF 27-OCT-1997; U19226.
 PR 01-NOV-1996; US-029930.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Black Mt, Hodgson JE, Knowles DGC, Lonetto MA, Nicholas RO,
 PI Reid RH, Zarfos PN;
 DR WPI: 98-286586/25.
 DR N-PSDB; V37409.
 PT New isolated nucleic acids from Streptococcus pneumoniae - useful,
 PT e.g. for identifying anti-bacterial(s) for treatment and prevention
 PT of meningitis
 PS Claim 11; Page 114; 130pp; English.
 CC The sequence is that of the polypeptide encoded by a region isolated
 CC from S. pneumoniae. The protein, or agonists of it,
 CC may be useful as an antibacterial for treatment or
 CC prevention of infection, specifically caused by S. pneumoniae
 CC (particularly meningitis) but possibly also Helicobacter
 CC pylori (ulcers and gastric cancer). It may be of particular
 CC use before insertion of an in-dwelling device or any other
 CC invasive procedure. The protein, or nucleic acid encoding
 CC it, can also be used in vaccines to induce a cellular
 CC and/or humoral immune response, or to screen for other
 CC antibacterials. The DNA may also contain flanking sequences
 CC that are potential sources of control elements for bacterial
 CC gene expression. Detecting a sequence encoding the protein
 CC can be used diagnostically, e.g. to detect a mutation for
 CC serotyping or classifying infectious agents.
 SQ Sequence 128 AA:

Query Match 30.28; Score 60; DB 33; Length 128;
 Best Local Similarity 53.88; Pred. No. 1.67e+02;
 Matches 7; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Db. 60 lkaerdifkrw-v 71
 | | : | | | | |
 OY 14 IKOVKKLEKKGW 26

Search completed: Wed Sep 15 10:38:07 1999
 Job time : 17 secs.

WILEY

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 15 10:40:03 1999; MasPar time 1.59 seconds
Tabular output not generated. 165.973 Million cell updates/sec

Title: >US-09-068-507A-1
Description: (1-26) from US09068507A.pep
Perfect Score: 199
Sequence: 1 KSSAYSLQMGATAIKVKKLFKRWG 26

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 21.296; Variance 76.695; scale 0.278

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	71	35.7	572	1	US-08-160- Sequence 3, Applicatio	8.35e+00
2	71	35.7	1620	2	US-08-542- Sequence 2, Applicatio	8.35e+00
3	60	30.2	572	2	US-08-185- Sequence 81, Applicati	8.17e+01
4	60	30.2	572	2	US-08-191- Sequence 81, Applicati	8.17e+01
5	59	29.6	231	2	US-08-808- Sequence 33, Applicati	9.99e+01
6	59	29.6	316	3	PCT-US95-0 Sequence 2, Applicatio	9.99e+01
7	59	29.6	316	2	US-08-464- Sequence 2, Applicatio	9.99e+01
8	59	29.6	372	4	5514582-2 Patent No. 5514582.	9.99e+01
9	59	29.6	372	2	US-08-513- Sequence 2, Applicatio	9.99e+01
10	59	29.6	385	2	US-08-340- Sequence 2, Applicatio	9.99e+01
11	59	29.6	385	2	US-08-461- Sequence 2, Applicatio	9.99e+01
12	58	29.1	410	1	US-08-073- Sequence 16, Applicati	1.12e+02
13	57	28.6	49	2	US-08-456- Sequence 26, Applicati	1.49e+02
14	57	28.6	49	2	US-08-237- Sequence 26, Applicati	1.49e+02
15	57	28.6	450	3	PCT-US95-0 Sequence 7, Applicati	1.49e+02
16	57	28.6	485	1	US-07-991- Sequence 42, Applicati	1.49e+02
17	57	28.6	544	1	US-08-264- Sequence 7, Applicatio	1.49e+02
18	57	28.6	1167	1	US-08-485- Sequence 6, Applicatio	1.49e+02
19	57	28.6	1167	1	US-08-590- Sequence 6, Applicatio	1.49e+02
20	57	28.6	1168	1	US-08-620- Sequence 9, Applicatio	1.49e+02
21	57	28.6	1220	1	US-08-158- Sequence 43, Applicati	1.49e+02
22	57	28.6	1220	2	US-08-611- Sequence 43, Applicati	1.49e+02
23	57	28.6	1289	4	5281530-3 Patent No. 5281530.	1.49e+02

24	57	28.6	1289	4	5426049-4	Patent No. 5426049.	1.49e+02
25	57	28.6	1289	1	US-08-158- Sequence 4, Applicatio	Sequence 4, Applicatio	1.49e+02
26	57	28.6	1289	1	US-07-876- Sequence 4, Applicatio	Sequence 4, Applicatio	1.49e+02
27	57	28.6	1289	1	US-08-083- Sequence 4, Applicatio	Sequence 4, Applicatio	1.49e+02
28	57	28.6	1289	2	US-08-316- Sequence 4, Applicatio	Sequence 4, Applicatio	1.49e+02
29	57	28.6	1289	1	US-08-304- Sequence 4, Applicatio	Sequence 4, Applicatio	1.49e+02
30	57	28.6	1289	3	PCT-US92-0 Sequence 4, Applicatio	Sequence 4, Applicatio	1.49e+02
31	57	28.6	1289	1	US-07-675- Sequence 4, Applicatio	Sequence 4, Applicatio	1.49e+02
32	57	28.6	1289	2	US-08-611- Sequence 4, Applicatio	Sequence 4, Applicatio	1.49e+02
33	57	28.6	1385	4	5426049-1	Patent No. 5426049.	1.49e+02
34	57	28.6	1385	1	US-07-876- Sequence 2, Applicatio	Sequence 2, Applicatio	1.49e+02
35	57	28.6	1385	4	US-08-316- Sequence 2, Applicatio	Sequence 2, Applicatio	1.49e+02
36	57	28.6	1385	4	5281530-1	Patent No. 5281530.	1.49e+02
37	57	28.6	1385	1	US-07-675- Sequence 2, Applicatio	Sequence 2, Applicatio	1.49e+02
38	57	28.6	1385	2	US-08-611- Sequence 2, Applicatio	Sequence 2, Applicatio	1.49e+02
39	57	28.6	1385	1	US-08-063- Sequence 2, Applicatio	Sequence 2, Applicatio	1.49e+02
40	57	28.6	1385	1	US-08-158- Sequence 2, Applicatio	Sequence 2, Applicatio	1.49e+02
41	57	28.6	1385	3	PCT-US92-0 Sequence 2, Applicatio	Sequence 2, Applicatio	1.49e+02
42	57	28.6	1385	1	US-08-304- Sequence 2, Applicatio	Sequence 2, Applicatio	1.49e+02
43	56	28.1	119	2	US-08-340- Sequence 19, Applicati	Sequence 19, Applicati	1.81e+02
44	56	28.1	338	1	US-08-553- Sequence 16, Applicati	Sequence 16, Applicati	1.81e+02
45	56	28.1	1023	1	US-08-198- Sequence 4, Applicatio	Sequence 4, Applicatio	1.81e+02

ALIGNMENTS

RESULT 1
ID US-08-160-861-3 STANDARD; PRT; 572 AA.

XX AC xxxxxx

XX XX

XX DT

XX DE

XX XX

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

CC CC

Sequence 3, Application US/08160861

Sequence 3, Application US/08160861

Patent No. 5529295

GENERAL INFORMATION:

APPLICANT: MORRIS, STEPHAN W

TITLE OF INVENTION: NOVEL FUSION NUCLEIC ACID SEQUENCES AND

TITLE OF INVENTION: FUSION PROTEINS PRESENT IN HUMAN t(2:5) LYMPHOMA, METHO

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX

STREET: 1100 NEW YORK AVE NW SUITE 600

CITY: WASHINGTON

STATE: D.C.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/160.861

FILING DATE: 02-DEC-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MILLMAN, ROBERT A

REGISTRATION NUMBER: 36217

REFERENCE/DOCKET NUMBER: 0656.0400000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2678

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 572 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 572 AA: 62849 MW; 1722781 CN;

Query Match 35.7%; Score 71; DB 1; Length 572;

Best Local Similarity 52.4%; Pred. No. 8.35e+00;
Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 2 SLOEGATGHHSCPOAMKKWGW 22
||| |||:
QY 6 SLOWGATAIKOVKKLFKKWGW 26

RESULT 2
ID US-08-542-363-2 STANDARD; PRT; 1620 AA.

XX

AC

XX

DT

XX

DE Sequence 2, Application US/08542363

XX Sequence 2, Application US/08542363

CC Patent No. 5770421

CC GENERAL INFORMATION:

CC APPLICANT: Morris, Stephan W.

CC APPLICANT: Look, A. Thomas

CC TITLE OF INVENTION: ALK Protein Tyrosine Kinase/Receptor and

CC TITLE OF INVENTION: Ligands Thereof

CC NUMBER OF SEQUENCES: 43

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

CC STREET: 1100 New York Avenue, N.W., Suite 600

CC CITY: Washington

CC STATE: DC

CC COUNTRY: USA

CC ZIP: 20005

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: IBM PC compatible

CC SOFTWARE: PC-DOS/MS-DOS

CC OPERATING SYSTEM: Patent In Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/542,363

CC FILING DATE: 12-OCT-1995

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Fox, Samuel L.

CC REGISTRATION NUMBER: 30,353

CC REFERENCE/DOCKET NUMBER: 0656.0400001/SLF/GKT

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 202-371-2600

CC TELEFAX: 202-371-2540

CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 1620 amino acids

CC TYPE: amino acid

CC STRANDEDNESS:

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 1620 AA; 176416 MW; 13627521 CN;

Query Match 35.7%; Score 71; DB 2; Length 1620;

Best Local Similarity 52.4%; Pred. No. 8.35e+00;

Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 895 SLOEGATGHHSCPOAMKKWGW 915

||| |||:
QY 6 SLOWGATAIKOVKKLFKKWGW 26

RESULT 3

ID US-08-185-949B-81 STANDARD; PRT; 572 AA.

XX

AC

XX

DT

XX

DE Sequence 81, Application US/08185949B

XX

CC

CC Sequence 81, Application US/08185949B

CC Patent No. 5874279

CC GENERAL INFORMATION:

CC APPLICANT: Mark D. Cochran

CC APPLICANT: Richard D. Macdonald

CC TITLE OF INVENTION: Recombinant Infectious Bovine

CC TITLE OF INVENTION: Rhinotracheitis Virus

CC NUMBER OF SEQUENCES: 104

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: John P. White

CC STREET: 1185 Avenue of the Americas

CC CITY: New York

CC STATE: New York

CC COUNTRY: USA

CC ZIP: 10036

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM 330 466 DX2

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/185,949B

CC FILING DATE: 03-NOV-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: White, John P.

CC REGISTRATION NUMBER: 678

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (212) 278-0400

CC TELEFAX: (212) 278-0525

CC INFORMATION FOR SEQ ID NO: 81:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 572 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 572 AA; 64607 MW; 1762281 CN;

SQ

Query Match 30.2%; Score 60; DB 2; Length 572;

Best Local Similarity 30.0%; Pred. No. 8.17e+01;

Matches 6; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 432 LQLGVIDISYNNIRINWTW 451

||| | : : : : |
QY 7 LQMGATAIKOVKKLFKKWGW 26

RESULT 4

ID US-08-191-866D-81 STANDARD; PRT; 572 AA.

XX

AC

XX

DT

XX

DE

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Wed Sep '15 12:53:59 1999

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/191,866D
CC FILING DATE: 4 February 1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P.
CC REGISTRATION NUMBER: 28,678
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 278-0400
CC TELEFAX: (212) 391-0525
CC TELEX: 422523
CC INFORMATION FOR SEQ ID NO: 81:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 572 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 572 AA; 64607 MW; 1762281 CN;
CC
Query Match 30.2%; Score 60; DB 2; Length 572;
Best Local Similarity 30.0%; Pred. No. 8.17e+01;
Matches 6; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
Db 432 LQMGVIDISDYNINRINWTW 451
Qy- 7 LQMGATAIKOVKKLFKKKGW 26
||| | | : : : : | |
RESULT 5
ID US-08-808-550-33 STANDARD; PRT; 231 AA.
XX-xxxxxx
XX
XX
XX
XX
XX
XX
Sequence 33, Application US/08808550
Sequence 33, Application US/08808550
Patent No. 5871992
GENERAL INFORMATION:
CC APPLICANT: Teebor, George W.
CC APPLICANT: Hilbert, Timothy P.
CC TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND
CC TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
CC NUMBER OF SEQUENCES: 42
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: David A. Jackson, Esq.
CC STREET: 411 Hackensack Ave, Continental Plaza, 4th
CC STREET: Floor
CC CITY: Hackensack
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 07601
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/808,550
CC FILING DATE: 26-FEB-1997
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jackson Esq., David A.
CC REGISTRATION NUMBER: 26,742
CC REFERENCE/DOCKET NUMBER: 1049-1-001 N
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 201-487-5800

CC TELEFAX: 201-343-11684
CC INFORMATION FOR SEQ ID NO: 33:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 231 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: S. cerevisiae
CC SEQUENCE 231 AA; 25982 MW; 264148 CN;
CC
Query Match 29.6%; Score 59; DB 2; Length 231;
Best Local Similarity 45.5%; Pred. No. 9.99e+01;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Db 161 HVHRLCKMWNW 171
Qy 16 QVKKLFKKKGW 26
||| | | : : : : | |
RESULT 6
ID PCT-US95-01827A-2 STANDARD; PRT; 316 AA.
XX-xxxxxx
XX
XX
XX
XX
XX
XX
Sequence 2, Application PC/TUS9501827A
Sequence 2, Application PC/TUS9501827A
GENERAL INFORMATION:
CC APPLICANT: HE, ET AL.
CC TITLE OF INVENTION: Human Prostatic Specific Reductase
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
CC ADDRESSEE: CECCHI, STEWART & OLSTEIN
CC STREET: 6 BECKER FARM ROAD
CC CITY: ROSELAND
CC STATE: NEW JERSEY
CC COUNTRY: USA
CC ZIP: 07068
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5 INCH DISKETTE
CC COMPUTER: IBM PS/2
CC OPERATING SYSTEM: MS-DOS
CC SOFTWARE: WORD PERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/01827A
CC FILING DATE: Concurrently
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: FERRARO, GREGORY D.
CC REGISTRATION NUMBER: 36,134
CC REFERENCE/DOCKET NUMBER: 325800-228
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 201-994-1700
CC TELEFAX: 201-994-1744
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 316 AMINO ACIDS
CC TYPE: AMINO ACID
CC STRANDEDNESS:
CC TOPOLOGY: LINEAR
CC MOLECULE TYPE: PROTEIN
CC SEQUENCE 316 AA; 35444 MW; 517444 CN;
CC


```

CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 385 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 385 AA; 43743 MW; 775024 CN;
SQ      29.68; Score 59; DB 2; Length 385;
      Query Match
      Best Local Similarity 62.5%; Pred. No. 9.99e+01;
      Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      29 NIFKLGW 36
      ::||| |||
QY      19 KLFKKGW 26

RESULT 11
      385 AA.

```

APPLICATION NUMBER: 07/786149
FILING DATE: 31-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/315015
FILING DATE: 23-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 565D1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

RESULT 11 PRT; 385 AA.
ID US-08-461-592B-2 STANDARD:
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application US/08461592B
XX
Sequence 2, Application US/08461592B
Patent No. 5834425
CC CC
GENERAL INFORMATION:
CC APPLICANT: Tedder, Thomas F.
CC APPLICANT: Kansas, Geoffrey S.
CC TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS
CC TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESSES:
CC

C	INFORMATION:
C	SEQUENCE CHARACTERISTICS:
C	LENGTH: 372 amino acids
C	TYPE: amino acid
C	TOPOLOGY: linear
C	SEQUENCE 372 AA; 42209 MW; 729258 CN;
EQ	Query Match 29.6%; Score 59; DB 2; Length 372; Best Local Similarity 62.5%; Pred. No. 9.99e+01; Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Dd	16 NIEKLGNW 23 ::
Qy	19 KLFKKGNW 26
RESULT 10	STANDARD; PRT; 385 AA.
ID- US-08-340-539A-2	

TITLE OF INVENTION: BLOCKING ASSEMBLY
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
 STREET: Ten Post Office Square
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/461,592B
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/340,539
 FILING DATE: 16-NOV-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/008,459
 FILING DATE: 25-JAN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: James F. Haley, Jr.
 REGISTRATION NUMBER: 27,794
 REFERENCE/DOCKET NUMBER: CG-104
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 596-9000
 TELEFAX: (212) 596-9090
 TELEX: 14-8367
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 385 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE 385 AA; 43743 MW; 775024 CA;
 Query Match 29.6%; Score 59; DB 2; Length 385;
 Best Local Similarity 62.5%; Pred. No. 9.99e+01;

RESULT 10 STANDARD; PRT; 385 AA.
ID- US-08-340-539A-2
XX
AC AC
XX XX
XX XX
DT DT
XX -
Sequence 2, Application US/08340539A
DE
Sequence 2, Application US/08340539A
XX
Patent No. 5808025
CC
GENERAL INFORMATION:
CC
APPLICANT: Tedder, Thomas F.
CC
TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS
CC
TITLE OF INVENTION: CHIMERIC SELECTINS FOR COMPONENT SELECTIN FUNCTION
CC
TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
CC
NUMBER OF SEQUENCES: 28
CC
CORRESPONDENCE ADDRESS:
CC
ADDRESSER: FISH & NEAVE
CC
STREET: 1251 Avenue Of the Americas
CC
CITY: New York
CC
STATE: New York
CC
COUNTRY: USA
CC
ZIP: 10020
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS
CC
SOFTWARE: Patent In Release #1.0, Version #1.30
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: US/08/340.539A
CC
FILING DATE: 16-NOV-1994
CC
CLASSIFICATION: 514
CC
PRIOR APPLICATION DATA:
CC
APPLICATION NUMBER: US 08/008,459
CC
FILING DATE: 25-JAN-1993
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Gunnison, Jane
CC
REGISTRATION NUMBER: 38,479
CC
REFERENCE/DOCKET NUMBER: CG-104 CON
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: 212-596-9000
CC
TELEFAX: 212-596-9090
CC
INFORMATION FOR SEQ ID NO: 2:

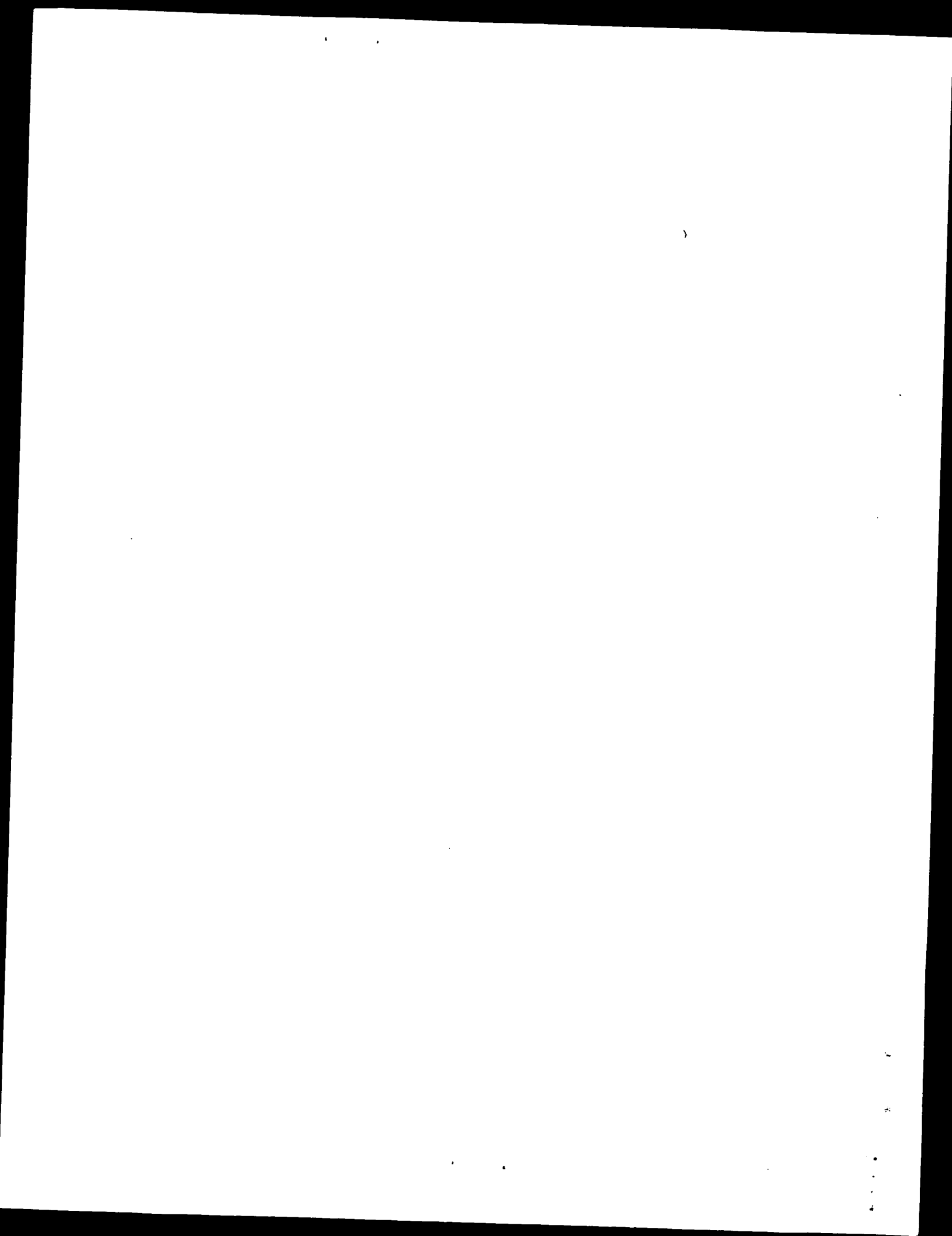

```
CC ZIP: 92037
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/237,401A
CC FILING DATE: 02-MAY-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/884,486
CC FILING DATE: 15-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Haile Ph.D., Lisa A.
CC REGISTRATION NUMBER: 38,347
CC REFERENCE/DOCKET NUMBER: 07251/007001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 678-5070
CC TELEFAX: (619) 678-5099
CC INFORMATION FOR SEQ ID NO: 26:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 49 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 49 AA; 5339 MW; 14108 CN;
CC
CC Query Match 28.68; Score 57; DB 2; Length
CC Best Local Similarity 33.38; Pred.No. 1.49e+02;
CC Matches 8; Conservative 6; Mismatches 10; In
DB 11 KVSDFGLTKAESQTDTGKLPVKW 34
QY 1 KSSAYSLOWGATAIKQVKKLFKKW 24
| | : | : | : | : | : | : |
| | : | : | : | : | : | : |
1 KSSAYSLOWGATAIKQVKKLFKKW 24

RESULT 15 STANDARD: PRT; 450 AA.
ID PCT-US95-05008-7
XX XXXXXX
AC
XX
XX
DT
XX
DE
XX
XX
CC Sequence 7, Application PC/TUS9505008
CC GENERAL INFORMATION:
CC APPLICANT: Sugen, Inc.
CC APPLICANT: 515 Galveston Drive
CC APPLICANT: Redwood City, California 94063-4720
CC APPLICANT: United States of America
CC APPLICANT: Wissenschaften E.V.
CC APPLICANT: Hofgarten Str. 2
CC APPLICANT: Munchen 80539
CC APPLICANT: Germany
CC TITLE OF INVENTION: Novel Megakaryocytic Prote
CC TITLE OF INVENTION: Kinases
CC NUMBER OF SEQUENCES: 21
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Pennie & Edmonds
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: U.S.A.
CC ZIP: 10036
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCI/US95/05008
```

```

CC FILING DATE: 24-APR-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC PRIOR APPLICATION NUMBER: US 08/232,545
CC APPLICATION NUMBER:
CC FILING DATE: 22-APR-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Coruzzi, Laura A.
CC REGISTRATION NUMBER: 30,742
CC REFERENCE/DOCKET NUMBER: 7683-074
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)790-9090
CC TELEFAX: (212)869-9741
CC TELEX: 66141 PENNIE
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 450 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: Protein
CC SEQUENCE 450 AA; 50704 MW, 1071733 CN;
SQ
Query Match 28.6%; Score 57; DB 3; Length 450;
Best Local Similarity 33.3; Pred. No. 1.49e+02;
Matches 8; Conservative 0; Gaps 0; Indels 0; Mismatches 10;
Db 329 KYSDFGLTKEASTQDTGKLPVKW 352
|||:::|:::|:::|:::|:::|
QY 1 KSSAYSLONGATAIKQVKKLFKKW 24

```



Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	Length				
1	166	83.4	22	2	A45913	plantaricin A - Lacto	5.88e+20
2	78	33.2	1609	2	S25345	probable membrane pro	7.06e-02
3	76	38.2	655	2	A42420	L-iduronidase (EC 3.2	1.56e-01
4	75	37.7	650	2	D71021	hypothetical protein	2.30e-01
5	75	37.7	960	2	JE0356	gamma-aminobutyric ac	2.30e-01
6	70	35.2	32	2	A40361	virC-region hypotheti	1.55e+00
7	69	34.7	649	2	B38129	bo-type ubiquinol oxi	2.25e+00
8	68	34.2	134	2	S27277	dedA protein - Mycoba	3.25e+00
9	68	34.2	157	2	C71060	hypothetical protein	3.25e+00
10	68	34.2	256	2	S41185	gene 38 protein - pha	4.68e+00
11	67	33.7	341	2	S31236	hypothetical protein	4.68e+00
12	67	33.7	457	2	A71174	hypothetical protein	4.68e+00
13	66	33.2	246	2	F69771	conserved hypothetica	6.70e+00
14	66	33.2	435	2	I52374	cerebrin-50 - human	6.70e+00
15	66	33.2	996	2	S42208	NAD+ ADP-ribosyltrans	6.70e+00
16	65	32.7	400	2	A37708	creatine kinase (EC 2	9.58e+00
17	64	32.2	529	2	I39841	transcription regulat	1.36e+01
18	64	32.2	835	2	F71638	hypothetical protein	1.36e+01
19	63	31.7	265	2	C70217	conserved hypothetica	1.93e+01
20	63	31.7	340	2	P02217	reverse transcriptase	1.93e+01
21	63	31.7	366	2	C69391	enolase (eno) homolog	1.93e+01
22	63	31.7	419	2	S42989	gene T48 protein - fr	1.93e+01
23	63	31.7	486	2	JC4240	lanosterol 14alpha-de	1.93e+01

Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

DNA Res. (1998) 5:35-76

Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.

#cross-references MUID:98344137

#accession D71021

##status preliminary; nucleic acid sequence not shown; translation not shown

##molecule_type DNA

##residues 1-650 ##label KAW

##cross-references GB:AP000008; NID:G326133; PID:di031515; PID:g3257899

##experimental_source strain OT3

##note this accession replaces an interim accession for a sequence replaced by GenBank

GENETICS

##gene

SUMMARY

PHI465

##length 650 ##molecular-weight 75553 ##checksum 3311

Query Match 37.7%; Score 75; DB 2; Length 650;

Best Local Similarity 58.3%; Pred. No. 2.30e-01;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 615 VKKIKIFKKG 626

QY 14 IKQVKLFKKG 25

RESULT 5

ENTRY JEO356

TITLE gamma-aminobutyric acid receptor B precursor - human

ALTERNATE_NAMES GABA(B) receptor

ORGANISM #Formal_name Homo sapiens #common_name man

DATE 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 12-Feb-1999

ACCESSIONS JEO356

REFERENCE JEO356

##authors Grifa, A.; Totaro, A.; Rommens, J.M.; Carella, M.; Roetto, A.; Borgato, L.; Zelante, L.; Gasparini, P.

##journal Biochem. Biophys. Res. Commun. (1998) 250:240-245

##title GABA (gamma-amino-butyric acid) neurotransmission: Identification and fine mapping of the human GABAB receptor gene.

##cross-references MUID:98440782

##accession JEO356

##molecule_type mRNA

##residues 1-960 ##label GRI

##cross-references GB:Y11044; NID:G2826760

##note this ORF is not annotated in GenBank entry HSGTHLAL, release 109

GENETICS

##map_position 6p21.3-6p21.3

KEYWORDS glycoprotein; neurotransmitter receptor; transmembrane protein

FEATURE

1-11

12-960

##domain signal sequence #status predicted #label SIG

##product gamma-aminobutyric acid receptor B #status predicted #label MAT

##domain transmembrane #status predicted #label TM1

##domain transmembrane #status predicted #label TM2

##domain transmembrane #status predicted #label TM3

##domain transmembrane #status predicted #label TM4

##domain transmembrane #status predicted #label TM5

##domain transmembrane #status predicted #label TM6

##domain transmembrane #status predicted #label TM7

##binding_site carbohydrate (Asn) (covalent) #status predicted

23,93,439,481,501,

513,630

US-09-068-507A-1.1pr

We Sep 15 12:54:01 1999

SUMMARY #length 960 #molecular-weight 108148 #checksum 3766

Query Match 37.7%; Score 75; DB 2; Length 960;

Best Local Similarity 87.5%; Pred. No. 2.30e-01; Indels 0; Gaps 0;

Matches 7; Conservative

Db 298 KLFKKGW 305

Qy 19 KLFKKGW 26

RESULT 6 A40361 #type complete

ENTRY virC-region hypothetical protein yscA - Versinia

TITLE enterocolitica plasmid pYV

ORGANISM #formal name Versinia enterocolitica

DATE 21-Feb-1992 #sequence_revision 21-Feb-1992 #text_change

09-Sep-1997

A40361

ACCESSIONS #accession A40361

REFERENCE #authors Michiels, T.; Vanooteghem, J.C.; Lambert de Rouvroit, C.;

#journal China, B.; Gustin, A.; Boudry, P.; Cornelis, G.R.

#title J. Bacteriol. (1991) 173:4994-5009

#cross-references Analysis of virC, an operon involved in the secretion of Yop

#accession proteins by Versinia enterocolitica.

#molecule_type A40361

#residues 1-32 #label MIC

#cross-references GB:M74011; NID:g155549; PID:g155550

GENETICS plasmid #length 32 #molecular-weight 3915 #checksum 9835

#genome 35.2%; Score 70; DB 2; Length 32;

SUMMARY Query Match 31.3%; Pred. No. 1.55e-00;

Best Local Similarity 31.3%; Pred. No. 1.55e-00;

Matches 5; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 1 MSQITTKHITVLFRRW 16

Qy 9 MGATAIKQVKLFKKW 24

RESULT 7 B38129 #type complete

ENTRY bo-type ubiquinol oxidase (EC 1.10.3.-) chain I - Bacillus

TITLE subtilis

ALTERNATE_NAMES cytochrome aa3 quinol oxidase chain I; quinol oxidase aa3-600

ORGANISM #formal name Bacillus subtilis

DATE 04-Mar-1993 #sequence_revision 15-Oct-1994 #text_change

17-Mar-1999

ACCESSIONS B38129; S39693; F69687

REFERENCE #authors Santana, M.; Kunst, F.; Hullo, M.F.; Rapoport, G.; Danchin,

#journal A.; Glaser, P.

#title J. Biol. Chem. (1992) 267:10225-10231

#cross-references Molecular cloning, sequencing, and physiological

#accession characterization of the qox operon from Bacillus subtilis

#molecule_type encoding the aa3-600 quinol oxidase.

#residues 1-649 #label SAN

#cross-references GB:M86548; NID:g143395; PID:g143397

#note sequence extracted from NCBI backbone (NCBIN:103632,

NCBIP:103603)

REFERENCE S39655

#authors Glaser, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales,

W.; Hullo, M.F.; Ionescu, M.; Lubochinsky, B.; Marcelino,

L.; Moszer, I.; Presecan, E.; Santana, M.; Schneider, E.;

Schweizer, J.; Vertes, A.; Rapoport, G.; Danchin, A.

#journal Mol. Microbiol. (1993) 10:371-384

#title

#cross-references MUID:95020537

#accession S39693

#molecule_type DNA

#residues 1-649 #label GLA

#cross-references EMBL:X73124; NID:g413923; PID:g413962

#accession A69580

REFERENCE

#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;

Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;

Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,

Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;

Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;

Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;

Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;

Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,

M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gollightly, E.J.; Grandi, G.;

S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;

Guiseppi, G.; Guy, B.J.; Hago, K.; Haefel, S.; Hosono, S.;

C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Karamata, D.;

Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Kobayashi,

Kashahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kumano, M.;

Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Lauber, J.;

Kurita, K.; Lapidus, A.; Lardinois, S.; Li, H.; Masuda, S.;

Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Mellado, R.P.;

Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.;

M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,

M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.;

V.; Pohl, T.M.; Portetle, D.; Portetle, B.; Prescott, G.;

A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;

Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;

Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;

Schleich, S.; Schroeter, R.; Scoffone, E.; Sekiguchi, J.;

Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,

B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Tanaka, T.;

Takemaru, K.; Takeuchi, M.; Takashi, A.; Uchiyama, S.;

Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;

Vandenberg, M.; Vannier, F.; Vassarotti, A.; Viari, A.;

Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;

Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,

K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;

Yoshikawa, H.; Danchin, A.

Nature (1997) 390:249-256

#journal

#title The complete genome sequence of the Gram-positive bacterium

#accession Bacillus subtilis.

#cross-references MUID:98044033

#status preliminary; nucleic acid sequence not shown;

#molecule_type DNA

#residues 1-649 #label KUN

#cross-references GB:299123; GB:AL009126; NID:g2636240; PID:el186315;

#experimental_source strain 168

GENETICS

#gene qoxB

#complex heterotetramer; chains I, II, III and IV

#classification #superfamily cytochrome-c oxidase chain I; cytochrome-c

#keywords chromoprotein; copper binding; electron transfer; heme;

heterotetramer; membrane-associated complex;

oxidoreductase; transmembrane protein

FEATURE

21-37

52-498

57-73

107-123

142-158

192-208

233-249

283-299

#domain transmembrane #status predicted #label TM1\

#domain cytochrome-c oxidase chain I homology #label

COL\

#domain transmembrane #status predicted #label TM2\

#domain transmembrane #status predicted #label TM3\

#domain transmembrane #status predicted #label TM4\

#domain transmembrane #status predicted #label TM5\

#domain transmembrane #status predicted #label TM6\

#domain transmembrane #status predicted #label TM7\

```

311-327      #domain transmembrane #status predicted #label TM8\
381-397      #domain transmembrane #status predicted #label TM9\
419-435      #domain transmembrane #status predicted #label TM9\
493-509      #domain transmembrane #status predicted #label TM10\
589-605      #domain transmembrane #status predicted #label TM11\
608-624      #domain transmembrane #status predicted #label TM12\
102-417      #binding_site heme a iron (His) (axial ligands) #status
280,329,330  #binding_site copper (His) #status predicted\
415          #binding_site heme a3 iron (His) (axial ligand) #status
SUMMARY      #length 649 #molecular-weight 73838 #checksum 999
Query Match  34.7%; Score 69; DB 2; Length 649;
Best Local Similarity 42.9%; Pred. No. 2.25e+00;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
Db 22 STALSTIAIIVLVYFKKWKW 42
QY 6 SLOWGATAIKOVKKLFRKKGW 26

RESULT 8
ENTRY   deda          #type complete
TITLE   B1177_C2_172 protein
ORGANISM #formal_name Mycobacterium leprae
DATE     19-Mar-1997 #sequence_revision 23-Apr-1997 #text_change
ACCESSIONS S72727
REFERENCE   S72693
#authors   Smith, D.R.; Robison, K.
#submission submitted to the EMBL Data Library, November 1993
#description Mycobacterium leprae cosmid B1177.
#accession  S72727
#status    Preliminary
#molecule_type DNA
#residues  1-134 #label SMI
GENETICS   ##cross-references EMBL:U00011; NID:g456807; PID:g466816
#gene      deda
SUMMARY    #length 134 #molecular-weight 14249 #checksum 9020
Query Match 34.2%; Score 68; DB 2; Length 134;
Best Local Similarity 28.8%; Pred. No. 3.25e+00;
Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;
Db 91 FPKHFGPGHVALVERLFNRWG 111
QY 5 YSLWGATAIKOVKKLFRKKGW 25

RESULT 9
ENTRY   C71060          #type complete
TITLE   hypothetical protein PH1177 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE     14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
ACCESSIONS C71060
REFERENCE   A71000
#authors   Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Negai, Y.; Sakai, M.; Ogura, K.; Otsuka, H.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. (1998) 5:55-76
#journal   Complete sequence and gene organization of the genome of a
#title     hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
OT3.
#cross-references MUID:98344137
#accession  C71060

```

```

##status      preliminary; nucleic acid sequence not shown;
##molecule_type DNA
##residues    1-157 #label KAW
##cross-references GB:AP000005; NID:g3236132; PID:dl031220; PID:g3257594
##experimental_source strain OT3
##note        this accession replaces an interim accession for a
sequence replaced by GenBank

GENETICS      PH1177
#gene         #length 157 #molecular-weight 18430 #checksum 3665
SUMMARY       Query Match 34.2%; Score 68; DB 2; Length 157;
Best Local Similarity 30.0%; Pred. No. 3.25e+00;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
Db 125 SLEIENKLTREERLYKKWS 144
QY 6 SLOWGATAIKOVKKLFRKKGW 25

RESULT 10
ENTRY   S41185          #type complete
TITLE   gene 38 protein - phage SPPI
ORGANISM #formal_name phage SPPI
DATE     20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
ACCESSIONS S43811; S41185
REFERENCE   S43798
#authors   Pedre, X.; Weise, F.; Chai, S.; Lueder, G.; Alonso, J.C.
#journal   J. Mol. Biol. (1994) 236:1324-1340
#title     Analysis of cis and trans acting elements required for the
initiation of DNA replication in the Bacillus subtilis
bacteriophage SPPI.
#cross-references MUID:94172631
#accession  S43811
#molecule_type DNA
#residues  1-256 #label PE2
GENETICS   ##cross-references EMBL:X67865; NID:g472886; PID:g439642
#gene      deda
SUMMARY    #length 256 #molecular-weight 29998 #checksum 6866
Query Match 34.2%; Score 68; DB 2; Length 256;
Best Local Similarity 38.5%; Pred. No. 3.25e+00;
Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
Db 60 ITSIRELMNRWG 72
QY 14 IKQVKLFRKKGW 26

RESULT 11
ENTRY   S31236          #type fragment
TITLE   hypothetical protein 1 (HIS3 3' region) - yeast
ORGANISM #formal_name Saccharomyces kluyveri (fragment)
DATE     18-Jun-1993 #sequence_revision 18-Jun-1993 #text_change
ACCESSIONS S31236; S23874
REFERENCE   S31234
#authors   Weinstein, K.G.; Strathern, J.N.
#journal   Yeast (1993) 9:351-361
#title     Molecular genetics in Saccharomyces kluyveri: the HIS3
homolog and its use as a selectable marker gene in S.
kluyveri and Saccharomyces cerevisiae.
#accession  S31236
#molecule_type DNA
#residues  1-341 #label WEI
GENETICS   ##cross-references EMBL:Z14125; NID:g4864; PID:g4867
#gene      deda
SUMMARY    #length 341 #checksum 822
Query Match 33.7%; Score 67; DB 2; Length 341;
Best Local Similarity 46.2%; Pred. No. 4.68e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```


DEPT. T 12

#journal DNA Res. (1998) 5:55-70
#title Complete sequence and gene organization of the genome of the hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.

```

#molecule_type DNA
#residues 1-457 #label KAW
#cross-references GB:AP000002; NID:g3236129; PID:di030621; PID:g3256995
#experimental_source strain OM3
#note this accession replaces an interim accession for a
#note this access replaced by GenBank

```

Query Match	33.7%	Score 0;	22	Gaps	0;
Best Local Similarity	41.7%	Pred. NO. 4.88e+00;		Indels	3;
Best Local Sim.	5.	Conservative	4;	Mismatches	4;

```

RESULT      13
ENTRY       F69771      #type complete
            conserved hypothetical protein ydbK - Bacillus subtilis
TITLE       #formal_name Bacillus subtilis
            #formal_name Bacillus subtilis
ORGANISM    05-Dec-1997 #sequence_revision
            05-Dec-1997 #sequence_revision
DATE        24-Sep-1998
            24-Sep-1998

```

F69771
 A69580
 #authors
 ACCESSIONS
 REFERENCE
 #authors

```
#cross-references MUID:98044033
#accession F69771
#status preliminary; nucleic acid sequence not shown;
#cannot not shown
```

```
##molecule_type DNA
##1-246 ##label KUN
##residues GB:299106; GB:AL009126; NID:g2632653; PID:e1192416;
##cross-References GB:299106; GB:AL009126; NID:g2632653; PID:e1192416;
##source strain 168
```

```

#experiment=1
GENETICS
#gene
ydbk      #length 246      #molecular-weight 27786  #checksum 1380
SUMMARY

```

Query Match	33.2%	Score 66;	DB 2e; 1;
Best Local Similarity	52.9%	Pred. No. 6.70e+00;	1; Gaps 1;
Conservative		6; Mismatches	1; Indels

175 GASTLVQLMKLFENKKG 191
10 GATAIKOVKKLF-KKKG 25

```

RESULT      14
ENTRY
ENTRY      152374      #type complete
ENTRY      cerebrin-50 - human
ENTRY      #formal_name Homo sapiens #common_name man
ENTRY      #title      02-Jul-1996 #text_change
ENTRY      ORGANISM
ENTRY      02-Jul-1996 #sequence_revision
ENTRY      DATE      02-Jul-1996

```

ACCESSIONS
I52374
I52374

I52374
I52374

Giacomelli, S.;
M.G.; Leone, P.

Li, A.H.; Cheng, C.Y. (1995) 35:135-144
 Blochem. Mol. Biol. Int. (1995) 35:135-144
 Cerebrin-50, a human cerebrospinal fluid protein whose mRNA
 is present in multiple tissues but predominantly expressed
 in the lymphoblastoid cells and the brain.
 in: *Proc. Soc. Neurosci.* 1995, 15:1076

```
#cross-references MUID:95253020
```

```

#accession
#status
#preliminary; translated from GB/EMBL/
I523/4

```

```

##molecule_type mRNA
##label RES
##1-435
##residues
PID:g913523
3563

```

```
#cross-references GB:S76853; NID:92101 #molecular-weight 51487  
#length 435 #checkssum 306
```

33.28;	Score 66;	DB 2;	Length 435;
	33.28;	Score 66;	5.70e+00;

Query Match	41.7%;	Pred. NO. 5.700	Indels
Best Local	10.0	4;	Mismatches 10;
conservative			

[illegible]

74 SVHSLOKNGEFMAZKE
| III : ::|| |
| III : ::|| |

26
| III : ::|| |
| III : ::|| |

3 SAYSLQMGATAIKQVANKELKRNQ

```

RESULT 15
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:94170813
#accession S42208
#molecule_type mRNA
##residues_type protein
##cross-references EMBL:D16482; NID:9473742; PID:d1004459; PID:9538248
#molecule_type protein
#residues 170-188:721-736;813-819;879-885
CLASSIFICATION #superfamily NAD+ ADP-ribosyltransferase
KEYWORDS DNA binding; glycosyltransferase; NAD; pentosyltransferase;
zinc finger
FEATURE
1-369 #domain DNA binding #status predicted #label DNA\
370-507 #domain auto-modification #status predicted #label NAD\
508-996 #domain NAD binding #status predicted #label AMO\
SUMMARY #length 996 #molecular-weight 113018 #checksum 1061
Query Match 33.2% Score 66; DB 2; Length 996;
Best Local Similarity 34.6% Pred.No. 6.70e+00;
Matches 9; Conservative 7; Mismatches 10; Indels 0; Gaps 0;
Db 552 KNSFYKQLLESMDKNREWFVSWGR 577
Oy 1 KSSAYSLOMGATAIKVKLFKKGW 26

```

Search completed: Wed Sep 15 10:38:41 1999
Job time : 17 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 15 10:38:57 1999; Maspar time 3.11 Seconds
Tabular output not generated. 236.414 Million cell updates/sec

Title: >US-09-068-507A-1
Description: (1-26) from US09068507A.pep
Perfect Score: 199
Sequence: 1 KSSAYSLMGATAIKQVKLFRKKGW 26

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 31.897; Variance 43.884; scale 0.727

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	199	100.0	48	1	PLNA_LACPL	5.90e-32
2	78	39.2	1609	1	FIG2_YEAST	1.20e-02
3	76	38.2	655	1	IDUA_CANFA	2.93e-02
4	70	35.2	32	1	YSCA_YEREN	3.91e-01
5	69	34.7	649	1	QOX1_BACSU	5.93e-01
6	68	34.2	134	1	Y007_MYCLE	8.97e-01
7	67	33.7	280	1	EAHA_HAFAL	1.35e+00
8	67	33.7	341	1	YHS3_SACKL	1.35e+00
9	66	33.2	996	1	PPOL_SARPE	2.02e+00
10	65	32.7	419	1	KCRS_CHICK	3.02e+00
11	64	32.2	529	1	YBBB_BACSU	4.48e+00
12	63	31.7	1376	1	VCAP_HSVEB	6.62e+00
13	62	31.2	102	1	YMI4_YEAST	9.73e+00
14	62	31.2	221	1	GTAL_RAT	9.73e+00
15	62	31.2	527	1	HSF8_LYCPE	9.73e+00
16	62	31.2	572	1	HMA_P13HW	9.73e+00
17	62	31.2	572	1	HMA_P13HU	9.73e+00
18	62	31.2	572	1	HMA_P13HU	9.73e+00
19	62	31.2	572	1	HMA_P13HV	9.73e+00
20	62	31.2	572	1	HMA_P13HT	9.73e+00
21	62	31.2	572	1	HMA_P13HX	9.73e+00
22	62	31.2	572	1	HMA_P13HX	9.73e+00
23	62	31.2	617	1	VC31_BPMD2	9.73e+00

ID	PLNA_LACPL	STANDARD;	PRT;	48 AA.		
24	62	31.2	901	1	MALT_ECOLI	9.73e+00
25	62	31.2	1131	1	YAB9_YEAST	9.73e+00
26	61	30.7	196	1	RETB_CHICK	1.42e+01
27	61	30.7	209	1	UBIG_SALTY	1.42e+01
28	61	30.7	218	1	YF07_MYCTU	1.42e+01
29	61	30.7	231	1	PF5_BACSU	1.42e+01
30	61	30.7	240	1	UBIG_ECOLI	1.42e+01
31	61	30.7	418	1	KCRU_MOUSE	1.42e+01
32	61	30.7	439	1	YGFP_ECOLI	1.42e+01
33	61	30.7	454	1	GLNA_METJA	1.42e+01
34	61	30.7	572	1	HEMA_P13B	1.42e+01
35	61	30.7	611	1	LEM2_CANFA	1.42e+01
36	61	30.7	724	1	VG01_BPP22	1.42e+01
37	61	30.7	766	1	DOC2_MOUSE	1.42e+01
38	61	30.7	770	1	DOC2_HUMAN	1.42e+01
39	61	30.7	1234	1	JM2_MOUSE	1.42e+01
40	60	30.2	170	1	Y531_METJA	2.07e+01
41	60	30.2	406	1	YC08_METJA	2.07e+01
42	60	30.2	466	1	DCOR_YEAST	2.07e+01
43	60	30.2	585	1	YEJM_HAEIN	2.07e+01
44	60	30.2	785	1	YE15_CAEEL	2.07e+01
45	60	30.2	1064	1	YY08_METJA	2.07e+01

ALIGNMENTS

RESULT 1
ID PLNA_LACPL STANDARD; PRT; 48 AA.
AC P80214;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE BACTERIOCIN PLANTARICIN A PRECURSOR.
GN PLNA.
OS LACTOBACILLUS PLANTARUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; LACTOBACILLACEAE;
OC LACTOBACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C11;
RX MEDLINE: 94161498.
RA DIEP B.D., HAVARSTEIN L.S., NISSEN-MEYER J., NES F.I.;
RT "The gene encoding plantaricin A, a bacteriocin from Lactobacillus plantarum C11, is located on the same transcription unit as an agr-like regulatory system";
RL APPL. ENVIRON. MICROBIOL. 60:160-166(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C11;
RX MEDLINE: 96345611.
RA DIEP D.B., HAVARSTEIN L.S., NES I.F.;
RT "Characterization of the locus responsible for the bacteriocin production in Lactobacillus plantarum C11";
RL J. BACTERIOL. 178:4472-4483(1996).
RN [3]
RP SEQUENCE OF 26-47.
RC STRAIN-C11;
RX MEDLINE: 94065628.
RA NISSEN-MEYER J., GRANLY LARSEN A.G., SLETTEN K., DAESCHEL M., NES I.F.;
RT "Purification and characterization of plantaricin A, a Lactobacillus plantarum bacteriocin whose activity depends on the action of two peptides";
RL J. GEN. MICROBIOL. 139:1973-1978(1993).
CC -!- FUNCTION: THIS HEAT STABLE BACTERIOCIN INHIBITS THE GROWTH OF CLOSELY RELATED LACTOBACILLUS SPECIES. IT MAY ACT AS A PORE-FORMING PROTEIN, CREATING A CHANNEL IN THE CELL MEMBRANE THROUGH A "BARREL STAVE" MECHANISM.
CC -!- SUBUNIT: ACTIVE PLANTARICIN A IS COMPOSED OF AN ALPHA CHAIN AND A BETA CHAIN
CC -!- THE BETA CHAIN SEQUENCE IS SHOWN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X75323; G452406; -;
 DR EMBL; X94434; E217591; -;
 DR PIR; A45913; A45913.
 KW ANTIBIOTIC; BACTERIOICIN.
 FT PROPEP 1 25
 FT CHAIN 26 48 BACTERIOICIN PLANTARICIN A.
 FT VARIANT 26 26 MISSING (IN THE ALPHA CHAIN).
 FT SEQUENCE 48 AA; 5458 MW; A6083DE9 CRC32;

Query Match 100.0%; Score 199; DB 1; Length 48;
 Best Local Similarity 100.0%; Pred. No. 5.90e-32;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 23 KSSAYSLQMGATAIKOVKLFKKWG 48
 QY 1 KSSAYSLQMGATAIKOVKLFKKWG 26

RESULT 2
 ID FIG2 YEAST STANDARD; PRT; 1609 AA.

AC P25653;
 DT 01-MAY-1992 (REL. 22, CREATED)

DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE FACTOR INDUCED GENE 2.

GN FIG2 OR YCR089W OR YCR89W OR YCR1102.

OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;

CC SACCHAROMYCETACEAE; SACCHAROMYCES.

CC [1]

CC SEQUENCE FROM N.A.

CC MEDLINE; 92397594.

CC WILSON C., GRISANTI P., FRONTALI L.;

CC "The complete sequence of a 6146 bp fragment of Saccharomyces
 CC cerevisiae chromosome III contains two new open reading frames.";

CC YEAST 8:569-575(1992).

CC -1- FUNCTION: REQUIRED FOR EFFICIENT MATING.

CC -1- INDUCTION: BY MATING PHEROMONES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; X59720; E264634; -;
 CC PIR; S19504; S19504.
 CC PIR; S25345; S25345.
 CC SGD; L000312; FIG2.

CC SEQUENCE 1609 AA; 166049 MW; DE974CE8 CRC32;

Query Match 39.2%; Score 78; DB 1; Length 1609;
 Best Local Similarity 34.8%; Pred. No. 1.20e-02;
 Matches 8; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 1092 PSOYSLSTATTINGIKTVYTTW 1114
 QY 2 SSAYSLQMGATAIKOVKLFKKW 24

RESULT 3
 ID IDUA CANFA STANDARD; PRT; 655 AA.
 AC Q01634;
 DT 01-FEB-1994 (REL. 28, CREATED)

DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE ALPHA-L-IDURONIDASE PRECURSOR (EC 3.2.1.76).
 GN IDUA.
 OS CANIS FAMILIARIS (DOG).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
 CC [1]
 CC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 CC TISSUE-TESTIS;
 CC MEDLINE; 92202199.
 RA STOLTZFUS L.J., SOSA-PINEDA B., MOSKOWITZ S.M., MENON K.P., DLOTT B.,
 RA HOOPER L., TEPLow D.B., SHULL R.M., NEUFELD E.F.;
 RT "Cloning and characterization of cDNA encoding canine alpha-L-
 RT iduronidase. mRNA deficiency in mucopolysaccharidosis I dog.";
 RL J. BIOL. CHEM. 267:6570-6575(1992).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP TISSUE-TESTIS, AND FIBROBLAST;
 RC MEDLINE; 93052413.
 RX MENON K.P., TIEU P.T., NEUFELD E.F.;

RT "Architecture of the canine IDUA gene and mutation underlying canine
 RT mucopolysaccharidosis I";
 RL GENOMICS 14:763-768(1992).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF ALPHA-L-IDURONOSIDIC LINKAGES IN
 CC DESULFATED DERMATAN.

CC -1- SUBUNIT: MONOMER (PROBABLE).

CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.

CC -1- TISSUE SPECIFICITY: FOUND UBQUITOUSLY.

CC -1- PTM: A SMALLER 63 KDA PROTEIN PROBABLY ARISES FROM IDUA PROTEIN
 CC BY PROTEOLYTIC CLEAVAGE.

CC -1- DISEASE: DEFECTS IN IDUA ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS
 CC TYPE I (MPS I).

CC -1- SIMILARITY: BELONGS TO FAMILY 39 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; L01058; -; NOT_ANNOTATED_CDS.

CC EMBL; L01059; -; NOT_ANNOTATED_CDS.

CC EMBL; L01060; -; NOT_ANNOTATED_CDS.

CC EMBL; L01061; -; NOT_ANNOTATED_CDS.

CC EMBL; L01065; G52348; -;

CC EMBL; M81893; G163964; -;

CC PIR; A42420; A42420.

CC PROSITE; PS01027; GLYCOSYL_HYDROL_F39; 1.

CC PFAM; PF01229; GLYCOSYL_hydr19; 1.

CC KW HYDROLASE; GLYCOSIDASE; LYSOSOME; SIGNAL.

CC SIGNAL 1 25

CC CHAIN 26 655 ALPHA-L-IDURONIDASE.

CC ACT_SITE 181 181 PROTON DONOR (POTENTIAL).

CC ACT_SITE 298 298 NUCLEOPHILE (POTENTIAL).

CC CARBOHYD 109 109 POTENTIAL.

CC CARBOHYD 189 189 POTENTIAL.

CC CARBOHYD 242 242 POTENTIAL.

CC CARBOHYD 335 335 POTENTIAL.

CC CARBOHYD 371 371 POTENTIAL.

CC SEQUENCE 655 AA; 72939 MW; 7358E9D7 CRC32;

Query Match 38.2%; Score 76; DB 1; Length 655;
 Best Local Similarity 33.3%; Pred. No. 2.93e-02;
 Matches 8; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

Db 267 SSIIYLEQEQATVQGIKRLFPKFA 290
 QY 2 SSAYSLQMGATAIKOVKLFKKWG 25

```

RESULT 4
ID YSCA_YEREN STANDARD; PRT; 32 AA.
AC Q01242;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE YOP PROTEINS TRANSLLOCATION PROTEIN A.
GN YSCA.
OS YERSINIA ENTEROCOLITICA.
OG PLASMID PYV.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC YERSINIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-439-80 / SEROTYPE O:9;
RC MEDLINE; 91317716.
RA MICHELIS T., VANOTEGHEM J.-C., DE ROUVROIT C., CHINA B., GUSTIN A.,
RT "Analysis of virC, an operon involved in the secretion of Yop
proteins by Yersinia enterocolitica.";
RL J. BACTERIOL. 173:4994-5009(1991).
CC -!- INDUCTION: AT 37 DEGREES CELSIUS IN THE ABSENCE OF CALCIUM
CC -!- BELONGS TO AN OPERON INVOLVED IN THE TRANSLLOCATION OF YOP PROTEINS
CC ACROSS THE BACTERIAL MEMBRANES OR IN THE SPECIFIC CONTROL OF THIS
CC FUNCTION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M74011; G155950; -
DR PIR; A40361; A40361.
KW PLASMID; VIRULENCE.
SQ SEQUENCE 32 AA; 3815 MW; 41DF6051 CRC32;

Query Match 35.2%; Score 70; DB 1; Length 32;
Best Local Similarity 31.3%; Pred. No. 3.91e-01;
Matches 5; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 1 MSQITTKHITVLFRRW 16
Qy 9 MGATAIKQVKLFKKW 24
| : |:: |::|
| : |:: |::|

RESULT 5
ID QOXL_BACSU STANDARD; PRT; 649 AA.
AC P34956;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE QUINOL OXIDASE POLYPEPTIDE I (EC 1.9.3.-) (QUINOL OXIDASE AA3-600,
DE SUBUNIT QOXB) (OXIDASE AA(3) SUBUNIT 1).
GN QOXB OR IPA-38D.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-168;
RC MEDLINE; 92268053.
RA SANTANA M., KUNST F., HULLO M.F., RAPOPORT G., DANCHIN A.,
RA GLASER P.;
RT "Molecular cloning, sequencing, and physiological characterization of
RT the qox operon from Bacillus subtilis encoding the aa3-600 quinol
RT oxidase.";
RL J. BIOL. CHEM. 267:10225-10231(1992).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC RX STRAIN-168;
RA GLASER P., KUNST F., ARNAUD M., COUDART M.P., GONZALES W.,
RA HULLO M.F., IONESCU M., LUBOCHINSKY B., MARCELINO L., MOSZER I.,
RA PRESECAN E., SANTANA M., SCHNEIDER E., SCHWEIZER J., VERTES A.,
RA RAPOPORT G., DANCHIN A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL MOL. MICROBIOL. 10:371-384(1993).
CC -!- FUNCTION: CATALYZES QUINOL OXIDATION WITH THE CONCOMITANT
CC REDUCTION OF OXYGEN TO WATER. MAJOR COMPONENT FOR ENERGY
CC CONVERSION DURING VEGETATIVE GROWTH.
CC -!- COFACTOR: HEMES A, A3, AND COPPER B.
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86548; G143397; -
DR EMBL; X73124; G413962; -
DR EMBL; Z99123; E1186315; -
DR PIR; B38129; B38129.
DR PIR; S39693; S39693.
DR SUBTILIST; BG10584; QOXB.
DR PROSITE; PS00077; COX1; 1.
DR PFAM; PF00115; COX1; 1.
DR HSP; P98002; 1A1.
KW OXIDOREDUCTASE; HEME; COPPER; TRANSMEMBRANE; RESPIRATORY CHAIN;
KW HYDROGEN ION TRANSPORT.
FT DOMAIN 1 15 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 16 34 POTENTIAL.
FT DOMAIN 35 56 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 57 75 POTENTIAL.
FT DOMAIN 76 97 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 98 117 POTENTIAL.
FT DOMAIN 118 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 157 POTENTIAL.
FT DOMAIN 158 190 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 191 209 POTENTIAL.
FT DOMAIN 210 227 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 228 246 POTENTIAL.
FT DOMAIN 247 272 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 273 292 POTENTIAL.
FT DOMAIN 293 315 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 316 335 POTENTIAL.
FT DOMAIN 336 343 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 344 362 POTENTIAL.
FT DOMAIN 363 377 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 378 397 POTENTIAL.
FT DOMAIN 398 405 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 406 425 POTENTIAL.
FT DOMAIN 426 452 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 453 472 POTENTIAL.
FT DOMAIN 473 490 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 491 510 POTENTIAL.
FT DOMAIN 511 584 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 585 604 POTENTIAL.
FT DOMAIN 605 610 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 611 631 POTENTIAL.
FT DOMAIN 632 649 CYTOPLASMIC (POTENTIAL).
FT METAL 102 102 IRON (HEME A) (PROBABLE).
FT METAL 280 280 COPPER B (PROBABLE).
FT METAL 284 284 COPPER B (PROBABLE).
FT METAL 329 329 COPPER B (PROBABLE).
FT METAL 330 330 COPPER B (PROBABLE).
FT METAL 415 415 IRON (HEME A3) (PROBABLE).

```

```

FT METAL 417 417 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 649 AA; 73838 MW; 7C64B76E CRC32;

Query Match 34.7%; Score 69; DB 1; Length 649;
Best Local Similarity 42.9%; Pred. No. 5.93e-01;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 22 STALSTIAIIFVLTVEFKKWK 42
   [ : : : : | | | | |
QY 6 SLQMGATAIKQVKLFFKKGW 26

RESULT 6
ID YQ07_MYCLE STANDARD; PRT; 134 AA.
AC Q49642;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 14.2 KD PROTEIN BL177_C2_172.
GN BL177_C2_172.
OS MYCOBACTERIUM LEPRAE.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA ROBISON K., SMITH D.R.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE DEDA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U00011; G466816;
KW HYPOTHETICAL PROTEIN
SQ SEQUENCE 134 AA; 14249 MW; 7F516A45 CRC32;

Query Match 34.2%; Score 68; DB 1; Length 134;
Best Local Similarity 28.6%; Pred. No. 8.97e-01;
Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 91 FPKHFGPGHVALVERLFNKG 111
   [ : : : : | : : : :
QY 5 YSLQMGATAIKQVKLFFKKGW 25

RESULT 7
ID EAEA_HAFAL STANDARD; PRT; 280 AA.
AC P52869;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE INTIMIN (OUTER MEMBRANE PROTEIN) (ATTACHING AND EFFACING PROTEIN)
DE (FRAGMENT).
GN EAEA OR EAE.
OS HAFNIA ALVEL.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC HAFNIA.
RN [1]
RP SEQUENCE FROM N.A.
RA FRANKEL G., CANDY D.C.A., EVEREST P., DOUGAN G.;
RL "Characterization of the C-terminal domains of intimin-like proteins
RT of enteropathogenic and enterohemorrhagic Escherichia coli,
RT Citrobacter freundii, and Hafnia alvei."
RL INFECT. IMMUN. 62:1835-1842(1994).
CC -!- FUNCTION: NECESSARY FOR THE PRODUCTION OF ATTACHING AND EFFACING
CC LESIONS ON TISSUE CULTURE CELLS.
CC -!- SUBCELLULAR LOCATION: OUTER SURFACE.

```

```

CC -!- SIMILARITY: BELONGS TO THE EAE/INVASIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L29509; G472359;
KW OUTER MEMBRANE.
FT NON_TER 1
SQ SEQUENCE 280 AA; 30146 MW; FBC2C19D CRC32;

Query Match 33.7%; Score 67; DB 1; Length 280;
Best Local Similarity 40.0%; Pred. No. 1.35e+00;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 209 PSSIKELKDLDDWG 223
   [ : : : : | : : : :
QY 11 ATAIKQVKLFFKKGW 25

RESULT 8
ID YHS3_SACKL STANDARD; PRT; 341 AA.
AC Q03000;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN IN HIS3 3' REGION (FRAGMENT).
OS SACCHAROMYCES KLUYVERI (YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CBS 3082;
RX MEDLINE; 93289813.
RA WEINSTOCK K.G., STRATHERN J.N.;
RT "Molecular genetics in Saccharomyces kluyveri: the HIS3 homolog and
RT its use as a selectable marker gene in S. kluyveri and Saccharomyces
RT cerevisiae."
RL YEAST 9:351-361(1993).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z14125; E1175791;
DR PIR; S31236; S31236.
KW HYPOTHETICAL PROTEIN.
FT NON_TER 1
FT DOMAIN 103 107 POLY-SER.
SQ SEQUENCE 341 AA; 37381 MW; 64772A58 CRC32;

Query Match 33.7%; Score 67; DB 1; Length 341;
Best Local Similarity 46.2%; Pred. No. 1.35e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 133 SQVSKLRKRVFKW 145
   [ : : : : | : : : :
QY 12 TAIKQVKLFFKKGW 24

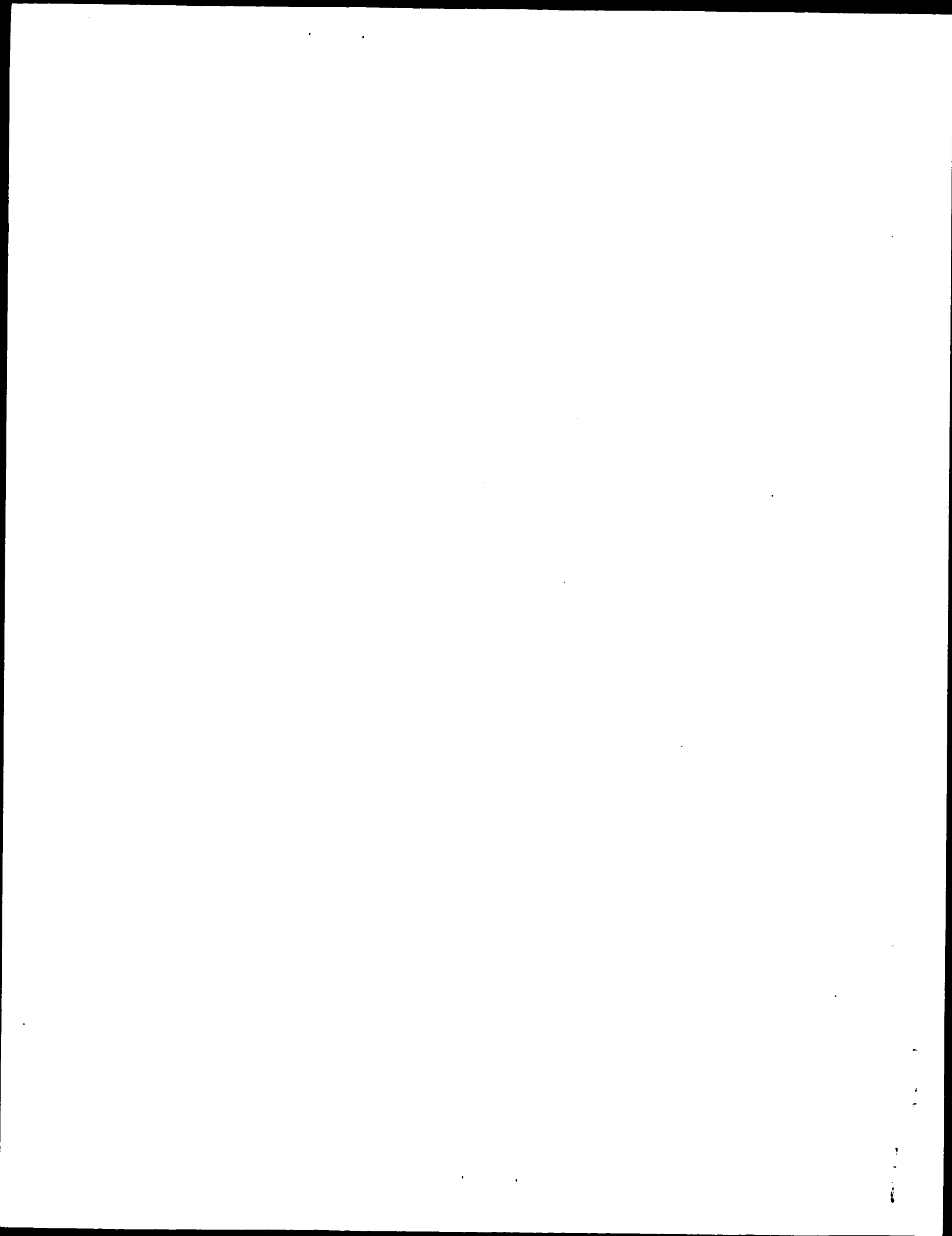
RESULT 9
ID PPOL_SARPE STANDARD; PRT; 996 AA.
AC Q11208;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

```


(CLASS-ALPHA) (CLONES PGST94 & PGTR261).
 DE RATTUS NORVEGICUS (RAT).
 OS EUKARYOTA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 84185691.
 RX LAI H.-C.J., LI N.-Q., WEISS M.J., REDDY C.C., TU C.-P.D.;
 RA "The nucleotide sequence of a rat liver glutathione S-transferase
 RT subunit cDNA clone.";
 RL J. BIOL. CHEM. 259:5536-5542(1984).
 [2]
 RN SEQUENCE OF 45-196 FROM N.A.
 RP MEDLINE: 82075944.
 RX KALINAK J.E., TAYLOR J.M.;
 RA "Rat glutathione S-transferase. Cloning of double-stranded cDNA and
 RT induction of its mRNA.";
 RL J. BIOL. CHEM. 257:523-530(1982).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-G.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- RAT LIVER GLUTATHIONE S-TRANSFERASES (GST) CONSIST OF BINARY
 CC COMBINATIONS OF THREE MAJOR CLASSES OF SUBUNITS DESIGNATED:
 CC YA (25KD), YB (27KD) AND YC (28KD).
 CC -!- YA AND YC SUBUNITS ARE ENCODED BY SEPARATE GENE FAMILIES OR
 CC TRANSCRIPTIONAL UNITS. THERE IS EXTENSIVE HOMOLOGIES BETWEEN
 CC THOSE TWO FAMILY OF SEQUENCES.
 CC -!- THE CLONES PGTR112 AND PGTR38 (SEE SEPARATE ENTRY) AND CLONES
 CC PGST94 AND PGTR261 WERE ISOLATED FROM DIFFERENT LITTERS OF
 CC SPRAGUE-DAWLEY RATS SUGGESTING THAT THESE ARE EITHER DIFFERENT
 CC ALLELES FOR THE SAME CATALYTIC FUNCTIONS OR DIFFERENT MEMBERS OF
 CC A CLOSELY RELATED MULTIGENE FAMILY.
 CC -!- IN ADDITION TO ITS ENZYMIC ACTIVITY, THE HOMODIMER OF YA CHAINS,
 CC CALLED LIGANDIN, BINDS VARIOUS ORGANIC ANIONS, XENOBIOTICS, AND
 CC AZOCARCINOGEN DYES. IT IS A CYTOSOLIC PROTEIN FOUND IN MANY
 CC MAMMALIAN TISSUES.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: K01931; G204495; -.
 DR PIR: A00591; XURTG.
 DR PFAM: PF00043; Gluts: 1.
 DR HSP: P08263; 1GSF.
 KW TRANSFERASE; MULTIGENE FAMILY.
 FT INIT_MET 0
 FT CONFLICT 151 151 K -> R (IN PIR DATA BANK).
 SQ SEQUENCE 221 AA; 25480 MW; C115D4C3 CRC32;
 Query Match 31.2%; Score 62; DB 1; Length 221;
 Best Local Similarity 43.8%; Pred. No. 9.73e+00;
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 Db 205 LPMDAKQIEEARKIFK 220
 QY 7 LQMGATAIKQVKLFK 22
 RESULT 15
 ID HSF8_LYCPE STANDARD; PRT: 527 AA.
 AC P41153;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE HEAT SHOCK FACTOR PROTEIN HSF8 (HEAT SHOCK TRANSCRIPTION FACTOR 8)
 DE (HSF8) (HEAT STRESS TRANSCRIPTION FACTOR).

GN HSF8.
 OS LYCOPERSICON PERUVIANUM (PERUVIAN TOMATO).
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
 OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANUM.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 94105354.
 RX SCHARF K.D., ROSE S., THIERFELDER J., NOVER L.;
 RA "Two cDNAs for tomato heat stress transcription factors.";
 RL PLANT PHYSIOL. 102:1355-1356(1993).
 CC -!- FUNCTION: DNA-BINDING PROTEIN THAT SPECIFICALLY BINDS HEAT SHOCK
 CC PROMOTER ELEMENTS (HSE) AND ACTIVATES TRANSCRIPTION (BY
 CC SIMILARITY).
 CC -!- SUBUNIT: HOMOTRIMER.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- PTM: EXHIBITS TEMPERATURE-DEPENDENT PHOSPHORYLATION
 CC (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE HSF FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X67600; G19492; -.
 DR PIR: S25481; S25481.
 DR PROSITE: PS00434; HSF_DOMAIN; 1.
 DR PFAM: PF00447; HSF_DNA-BIND; 1.
 DR HSP: P21121; 3HSF.
 DR TRANSFAC: T02033; -.
 KW TRANSCRIPTION REGULATION; NUCLEAR PROTEIN; DNA-BINDING; ACTIVATOR;
 KW PHOSPHORYLATION; HEAT SHOCK; MULTIGENE FAMILY.
 FT DNA_BIND 39 133 BY SIMILARITY.
 FT DOMAIN 14 20 POLY-GLY.
 FT DOMAIN 139 143 POLY-GLN.
 FT DOMAIN 195 198 POLY-GLN.
 SQ SEQUENCE 527 AA; 57519 MW; 34DE58C9 CRC32;
 Query Match 31.2%; Score 62; DB 1; Length 527;
 Best Local Similarity 64.3%; Pred. No. 9.73e+00;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Db 514 QMGATNKKQKHMI 527
 QY 8 QMGATAIKQVKLF 21

Search completed: Wed Sep 15 10:39:05 1999
 Job time : 8 secs.



[illegible]

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.

```

mpsrch_pp      protein - protein database search, using Smith-Waterman algorithm
               Wed Sep 15 10:39:22 1999;   Maspar time 5.79 Seconds
               Run on:                      245.104 Million cell updates/sec
               Regular output not generated.

```

```

Title:
Description:
Perfect Score:
sequence:
>US-09-068-507A-1
(1-26) from US09068507A.pap
199
1 KSSAYSLONGATAIKQVKKLEFKKNGW 26

```

Scoring table: PAM 150
Gap 15

179066 seqs, 54579741 residues
searched.

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:
sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus
```

Statistics: Mean 30.932; Variance 45.599; scale 0.678

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, σ^2 is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	77	38.7	822	11	054923	RSEC15.	5.45e-02
2	75	37.7	850	11	059134	550AA LONG HYPOTHETICA	1.27e-01
3	75	37.7	844	11	008621	GABA-BR1B RECEPTOR.	1.27e-01
4	75	37.7	960	11	008520	GABA-BR1A RECEPTOR.	1.27e-01
5	73	36.7	950	14	Q66578	GLYCOPROTEIN GP14.	2.92e-01
6	72	36.2	218	5	Q17690	COSMID C06A8.	4.40e-01
7	69	34.7	450	13	073786	C-SRC KINASE.	1.48e+00
8	69	34.7	1170	10	048573	PUTATIVE DISEASE RESIS	1.48e+00
9	68	34.2	157	1	058919	157AA LONG HYPOTHETICA	2.20e+00
10	68	34.2	214	2	049635	HYPOTHETICAL 23.1 KD P	2.20e+00
11	68	34.2	256	9	Q38150	PHAGE SP1 DNA SEQUENC	2.20e+00
12	67	33.7	51	13	Q31567	PROTEIN-TYROSINE KINAS	3.26e+00
13	67	33.7	225	14	Q66072	VIRION PROTEIN (P3ACME	3.26e+00
14	67	33.7	457	1	058328	457AA LONG HYPOTHETICA	3.26e+00
15	67	33.7	646	14	Q37319	NUCLEOSIDE TRIPHOSPHAT	3.26e+00
16	67	33.7	912	5	Q19712	F2B5.8.	3.26e+00
17	67	33.7	1415	5	Q61442	ZYG-9.	3.26e+00
18	66	33.2	246	2	P96606	PROBABLE MEMBRANE SPAN	4.80e+00
19	66	33.2	375	4	Q15150	MYELIN TRANSCRIPTION F	4.80e+00
20	66	33.2	435	4	Q16301	CEREBRIN-50.	4.80e+00

21	66	33.2	767	5	Q20176	COSMID F38E9.	4.80e+00
22	65	32.7	397	5	Q16371	F29G9.6 PROTEIN.	7.03e+00
23	65	32.7	492	14	Q80624	GAG PROTEIN.	7.03e+00
24	64	32.2	278	2	Q86433	UDP-GLUCOSE PYROPHOSPH	1.03e+01
25	64	32.2	370	5	Q86719	F22E5.1 PROTEIN.	1.03e+01
26	64	32.2	370	5	Q86719	POLLEN-SPECIFIC 1-AMIN	1.03e+01
27	64	32.2	470	10	Q65028	H19N07.2B PROTEIN.	1.03e+01
28	64	32.2	1133	5	Q45624	H19N07.2B PROTEIN.	1.03e+01
29	63	31.7	176	5	Q26633	E-ABL MRNA (FRAGMENT).	1.50e+01
30	63	31.7	230	14	Q88960	DNA POLYMERASE (FRAGME	1.50e+01
31	63	31.7	265	2	Q50975	CONSERVED HYPOTHETICAL	1.50e+01
32	63	31.7	276	11	Q63777	HYPOTHETICAL 32.0 KD P	1.50e+01
33	63	31.7	340	11	Q64175	REVERSE TRANSCRIPTASE	1.50e+01
34	63	31.7	366	1	Q29133	ENOLASE (ENO).	1.50e+01
35	63	31.7	367	10	Q40933	NTH1.	1.50e+01
36	63	31.7	387	2	Q08670	PUTATIVE SERINE PROTEA	1.50e+01
37	63	31.7	419	5	Q24548	T48 MRNA PRECURSOR.	1.50e+01
38	63	31.7	513	11	Q63289	L1 RETROPOSON ORF2 MR	1.50e+01
39	63	31.7	536	11	Q63305	LONG INTERSPERSED REPE	1.50e+01
40	63	31.7	559	2	Q44107	MAJOR SURFACE PROTEIN	1.50e+01
41	63	31.7	600	3	Q74577	HYPOTHETICAL 68.8 KD P	1.50e+01
42	63	31.7	646	11	Q63779	HYPOTHETICAL 76.2 KD P	1.50e+01
43	63	31.7	685	11	Q63288	L1 RETROPOSON, ORF2 MR	1.50e+01
44	63	31.7	1300	11	P97692	L1 RETROTRANSPONSON ORF	1.50e+01
45	62	31.2	574	14	Q89237	HEMAGGLUTININ-NEURAMIN	2.18e+01

ALIGNMENTS

RESULT	1
ID	O54923
AC	PRELIMINARY; PRT; 822 AA.
DT	01-JUN-1998 (TREMBUREL. 06, CREATED)
DT	01-JUN-1998 (TREMBUREL. 06, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBUREL. 08, LAST ANNOTATION UPDATE)
DE	RSEC15.
OS	RATTUS NORVEGICUS (RAT).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC	SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=BRAIN:
RX	MEDLINE; 9607070. HAZUKA C.D., PETERSON K.E., HSU S.C.,
RA	KEE Y., YOO J.S.,
RA	SCHELLER R.H.;
RL	"Subunit structure of the mammalian exocyst complex." ;
RL	PROC. NATL. ACAD. SCI. U.S.A. 94:14438-14443(1997).
DR	EMBL; AF032668; G2827162; -
SO	SEQUENCE 822 AA; 95105 MW; 38EFF544 CRC32;

```
Query Match      38.7%;   Score 77;   DB 11;   Length 822;
Best Local Similarity 33.3%;   Pred. No. 5.45e-02;
Matches 7;   Conservative 7;   Mismatches 7;   Indels 0;   Gaps 0;
```

D_b 731 TLQLAFIDLRQLLDLFMVWDW 751
 :||:: ::|: ||: | |
O_v 6 SIOMGATAIKOVKKLEFKKGGW 26

RESULT 2 PRELIMINARY: PRT: 650 AA.

AC O59134; (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DT 6503A LONG HYPOTHETICAL PROTEIN.

GN PYROCOCCUS HORIKOSHII.
OS ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCUS.
QC [1]
RN SEQUENCE FROM N.A.
RP STRAIN=OT3;
RC MEDLINE; 98344137.
FX

—

LEMBACH D.I.	TO EMBL/GENBANK/DBJ DATA BANKS.
SUBMITTED (NOV-1995)	
(3)	
SEQUENCE FROM N.A.	
WATERSTON R.I.	TO EMBL/GENBANK/DBJ DATA BANKS.
SUBMITTED (NOV-1995)	
EMBL: U93849; G1053042;	
DR	
SEQUENCE 218 AA; 25556 MW; 1ACF80C3 CRC32;	
Seq	
Query Match	36.2%; Score 72; DB 5; Length 218;
Best Local Similarity	60.0%; Pred. No. 4; 40e-01;
Matches	9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
DB	
106	IAATSVKQVEKLFQK 120
QY	9 MGATAIKQVKLFKK 23
RESULT 7	PRELIMINARY; PRT; 450 AA.
ID	073786
AC	073786;
DI	01-AUG-1998 (TREMBLREL. 07, CREATED)
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT	01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE	C-SRC KINASE.
GN	CSK.
OS	XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC	MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN	[1]
RP	SEQUENCE FROM N.A.
RA	MURPHY S.M., MORGAN D.O.;
RL	SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR	EMBL: AF052430; G2967840;
SQ	SEQUENCE 450 AA; 50807 MW; E9717367 CRC32;
	Query Match 34.7%; Score 69; DB 13; Length 450;
	Best Local Similarity 41.7%; Pred. No. 1; 48e+00;
	Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
DB	329 KVSDFGLTKESAIQDTSKLPVKW 352
QY	1 KSSAYSILQMGATAIKQVKLFKKW 24
RESULT 8	PRELIMINARY; PRT; 1170 AA.
ID	048573
AC	048573;
DT	01-JUN-1998 (TREMBLREL. 06, CREATED)
DT	01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT	01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE	PUTATIVE DISEASE RESISTANCE PROTEIN.
GN	T19K24.2.
OS	ARABIDOPSIS THALIANA (MOUSE-EAR CRESS);
OC	EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC	EUPHYLLOPHTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC	CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-CV. COLUMBIA;
RA	ROUNSEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.;
RA	SYRES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.;
RA	SOMERVILLE C.R., VENTER J.C.;
RL	SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR	EMBL: AC002342; G2660663;
SQ	SEQUENCE 1170 AA; 133397 MW; 9B32952B CRC32;
	Query Match 34.7%; Score 69; DB 10; Length 1170;
	Best Local Similarity 54.5%; Pred. No. 1; 48e+00;
	Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
DB	372 QIETLFGKWNW 382

Query Match
Best Local Similarity 34.2%; Score 68; DB 2; Length 214;
Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 91 FPKHFGHVALVERLNRWG 111
QY 5 YSLQMGATAIKOVKLFKKW 25

RESULT 11
ID Q38150 PRELIMINARY; PRT; 256 AA.
AC Q38150;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DE PHAGE SP1 DNA SEQUENCE CODING FOR PRODUCTS REQUIRED FOR REPLICATION
DE INITIATION.
GN 38.
OS BACTERIOPHAGE SP1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE;
OC LAMEDA PHAGE GROUP.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94172631.
RA PEDRE X., WEISE F., CHAI S., LUEDER G., ALONSO J.C.;
RT "Analysis of cis and trans acting elements required for the
RT initiation of DNA replication in the Bacillus subtilis bacteriophage
RT SP1.";
RL J. MOL. BIOL. 236:1324-1340(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA ALONSO J.C.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X67865; G439642; -.
DR EMBL; X97918; E244864; -.
SQ SEQUENCE 256 AA; 29998 MW; 16837A2C CRC32;

Query Match
Best Local Similarity 34.2%; Score 68; DB 9; Length 256;
Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 60 ITSIRELMNRWG 72
QY 14 IKQVKLFKKW 26

RESULT 12
ID Q91567 PRELIMINARY; PRT; 51 AA.
AC Q91567;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DE PROTEIN-TYROSINE KINASE (G56) (FRAGMENT).
GN G56.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-EMBRYO;
RA MEDLINE; 95383727.
RA BRANDLI A.W., KIRSCHNER M.W.;
RT "Molecular cloning of tyrosine kinases in the early Xenopus embryo:
RT identification of Eck-related genes expressed in cranial neural crest
RL cells of the second (hyoid) arch.";
RL DEV. DYN. 203:119-140(1995).
DR EMBL; U11729; G509579; -.
DR PFAM; PF00069; pkinase; 1.
KW TYROSINE-PROTEIN KINASE.

FT NON_TER 1 1
SQ NON_TER 51 51
SQ SEQUENCE 51 AA; 5577 MW; 84274507 CRC32;

Query Match
Best Local Similarity 33.7%; Score 67; DB 13; Length 51;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Db 13 KVSDFGLTKESAIQDSSKLPVKW 36
QY 1 KSSAYSLOMGATAIKOVKLFKKW 24

RESULT 13
ID Q66072 PRELIMINARY; PRT; 225 AA.
AC Q66072;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DE VIRION PROTEIN (FRAGMENT).
OS CANINE HERPESVIRUS.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-MILOU;
RA REMOND M., SHELDRICK P., LEBRETON F., NARDEUX P., FOULON T.;
RL J. VIROL. 77:37-48(1996).
DR EMBL; X90443; G1066229; -.
FT NON_TER 225 225
SQ SEQUENCE 225 AA; 25754 MW; E3E43F9B CRC32;

Query Match
Best Local Similarity 33.7%; Score 67; DB 14; Length 225;
Matches 8; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Db 111 AKVYGLQGEAKVADQLFTTW 133
QY 2 SSAYSLOMGATAIKOVKLFKKW 24

RESULT 14
ID Q58328 PRELIMINARY; PRT; 457 AA.
AC Q58328;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DE 457AA LONG HYPOTHETICAL PROTEIN.
GN PH0589.
OS PYROCOCCUS HORIKOSHII.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAWARABAVASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSIYAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete Sequence and Gene Organization of the Genome of a
RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 5:55-76(1998).
DR EMBL; AP000002; D1030621; -.
SQ SEQUENCE 457 AA; 52090 MW; CD2283F7 CRC32;

Query Match
Best Local Similarity 33.7%; Score 67; DB 1; Length 457;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 97 REVREYMRKKGW 108
QY 108

US-09-068-507A-1.ispt

Wed Sep 15 12:54:02 1999

OY 15 KOVKKLFKKW 26

RESULT 15 PRELIMINARY; PRT; 646 AA.
 ID O37319
 AC O37319; 1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DE NUCLEOSIDE TRIPHOSPHATASE I (EC 3.6.1.15) (NUCLEOSIDE-TRIPHOSPHATASE)
 DE (NUCLEOSIDE TRIPHOSPHATASE PHOSPHOHYDROLASE).
 DE (NUCLEOSIDE TRIPHOSPHATASE PHOSPHOHYDROLASE).
 DE HELIOTHIS ARMIGERA ENTOMOPOXVIRUS (HAEPV).
 OS VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; ENTOMOPOXVIRINAE.
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; ENTOMOPOXVIRINAE.
 RN SEQUENCE FROM N.A.
 RP STRAIN-CLONAL ISOLATE WT#2/011293;
 RC STRAIN-CLONAL ISOLATE WT#2/011293;
 RA SRISKANTHA A., OSBORNE R.J., DALL D.J.;
 RL J. GEN. VIROL. 0:0-0(1997).
 CC -1- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + PHOSPHATE.
 DR EMBL; AF019224; G2460172;
 DR PFAM; PF00176; SNF2_N; 1.
 DR PFAM; PF00271; helicase_C; 1.
 DR HYDROLASE.
 KW HYDROLASE.
 SQ SEQUENCE 646 AA; 75763 MW; 1C626ECF CRC32;

Query Match 33.7%; Score 67; DB 14; Length 646;
 Best Local Similarity 54.5%; Pred. No. 3.26e+00;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 75 IKHKELEPRW 85
 OY 14 IKQVKKLFKKW 24

Search completed: Wed Sep 15 10:39:47 1999
 Job time : 25 secs.

 W P S R E L A

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Sep 15 10:38:16 1999; MasPar time 262.06 Seconds
 Tabular output not generated. 1205.840 Million cell updates/sec

Title: >US-09-068-507A-2
 Description: (1-114) from US09068507A.seq
 Perfect Score: 114
 N.A. Sequence: 1 ATGATGATATTTAAACT.....AAATTTTACCATCGTTAA 114
 Comp: TACTACTATAAATTTTGA.....TTTAAATGGGTAGCAAT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: emb158
 1:em_bal 2:em_bal 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
 7:em_in 8:em_om 9:em_or 10:em_ov 11:em_pat 12:em_ph
 13:em_pl 14:em_ro 15:em_sts 16:em_vl
 Database: genbank111
 17:gb_bal 18:gb_bal 19:gb_htg1 20:gb_htg2 21:gb_in1
 22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
 28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
 33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vl

Statistics: Mean 9.118; Variance 6.864; scale 1.328

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	114	100.0	5873	17	Lactobacillus sake ind	7.64e-44
2	114	100.0	7597	17	LSSAKACNP	7.54e-44
3	31	27.2	10772	21	Drosophila melanogaste	3.81e-02
4	29	25.4	216021	31	Homo sapiens Chromosom	2.65e-01
5	29	25.4	216021	31	HUAC004787	2.65e-01
6	27	23.7	7218	25	Sequence 14 from paten	1.76e-00
7	26	22.8	1056	23	Mustela vison GT dinuc	4.43e+00
8	26	22.8	7430	28	Camptotheca acuminata	4.43e+00
9	26	22.8	10772	21	Drosophila melanogaste	4.43e+00
10	26	22.8	31933	22	Caenorhabditis elegans	4.43e+00
11	26	22.8	38856	21	Caenorhabditis elegans	4.43e+00
12	26	22.8	44996	21	Caenorhabditis elegans	4.43e+00
13	26	22.8	91638	19	Homo sapiens chromosom	4.43e+00

c	14	26	22.8	158078	31	AC005094	Homo sapiens BAC clone	4.43e+00
c	15	26	22.8	198526	20	AC006280	Plasmodium falciparum	4.43e+00
c	16	25	21.9	1891	21	DDU73685	Dictyostelium discoideum	1.10e+01
c	17	25	21.9	3108	27	LELDHIGN	Lycopersicon esculentum	1.10e+01
c	18	25	21.9	6993	29	AB006625	Homo sapiens mRNA for	1.10e+01
c	19	25	21.9	20962	21	CEY48E1C	Caenorhabditis elegans	1.10e+01
c	20	25	21.9	28220	21	CEM176	Caenorhabditis elegans	1.10e+01
c	21	25	21.9	38225	21	CEY13G11	Caenorhabditis elegans	1.10e+01
c	22	25	21.9	73210	19	AC005653	Drosophila melanogaster	1.10e+01
c	23	25	21.9	80353	20	AC004710	Plasmodium falciparum	1.10e+01
c	24	25	21.9	178367	31	AC006115	Homo sapiens chromosome	1.10e+01
c	25	25	21.9	191908	19	CEY51B9	Caenorhabditis elegans	1.10e+01
c	26	25	21.9	291821	19	CEY48E1	Caenorhabditis elegans	1.10e+01
c	27	24	21.1	868	28	AF006509	Endogone pisiformis 18	2.69e+01
c	28	24	21.1	871	28	AF006510	Endogone pisiformis 18	2.69e+01
c	29	24	21.1	878	28	AF006511	Endogone pisiformis 18	2.69e+01
c	30	24	21.1	25359	21	CEY02D10	Caenorhabditis elegans	2.69e+01
c	31	24	21.1	30563	21	CEY27F6	Caenorhabditis elegans	2.69e+01
c	32	24	21.1	39356	21	CEY32A11	Caenorhabditis elegans	2.69e+01
c	33	24	21.1	42696	21	CELC45G7	Caenorhabditis elegans	2.69e+01
c	34	24	21.1	48338	19	CEY02D9	Caenorhabditis elegans	2.69e+01
c	35	24	21.1	58663	22	AC004300	Drosophila melanogaster	2.69e+01
c	36	24	21.1	74371	31	AC005369	Homo sapiens chromosome	2.69e+01
c	37	24	21.1	82966	19	CEAA2	Caenorhabditis elegans	2.69e+01
c	38	24	21.1	120011	19	HS124K22	Human DNA sequence ***	2.69e+01
c	39	24	21.1	137678	29	HSY79K22	Human DNA sequence from	2.69e+01
c	40	24	21.1	182152	19	CEY67A10	Caenorhabditis elegans	2.69e+01
c	41	24	21.1	183422	19	AC006905	Caenorhabditis elegans	2.69e+01
c	42	24	21.1	195349	19	AC006705	Caenorhabditis elegans	2.69e+01
c	43	24	21.1	228418	20	AC005505	Plasmodium falciparum	2.69e+01
c	44	24	21.1	257817	19	AC006909	Caenorhabditis elegans	2.69e+01
c	45	24	21.1	299308	19	AC006898	Caenorhabditis elegans	2.69e+01

ALIGNMENTS

RESULT	1	AF002276	5873 bp	DNA	BCT	01-JAN-1998
LOCUS		Lactobacillus sake inducing peptide preprotein (sprip), histidine				
DEFINITION		protein kinase homolog (sppk), response regulator homolog (sppr),				
		sakacin P preprotein (sppA), putative sakacin P immunity protein				
		(sppA) and putative ATP-dependent translocator protein (sppT)				
		genes, complete cds.				
ACCESSION		AF002276				
NID		92735685				
VERSION		AF002276.1	GI:2735685			
KEYWORDS		Lactobacillus sake.				
SOURCE		Lactobacillus sake.				
ORGANISM		Eubacteria; Firmicutes; Low G+C gram-positive bacteria;				
		Lactobacillaceae; Lactobacillus.				
REFERENCE		1 (bases 1 to 5873)				
AUTHORS		Eijsink V.G., Brurberg, M.B., Middelhoven, P.H. and Nes, I.F.				
TITLE		Induction of bacteriocin production in Lactobacillus sake by a				
		secreted peptide				
JOURNAL		J. Bacteriol. 178 (8), 2232-2237 (1996)				
MEDLINE		96218690				
REFERENCE		2 (bases 1 to 5873)				
AUTHORS		Brurberg, M.B., Nes, I.F. and Eijsink, V.G.				
TITLE		Pheromone-induced production of antimicrobial peptides in				
		Lactobacillus				
JOURNAL		Mol. Microbiol. 26 (2), 347-360 (1997)				
MEDLINE		98043507				
REFERENCE		3 (bases 1 to 5873)				
AUTHORS		Brurberg, M.B. and Eijsink, V.G.H.				
TITLE		Direct Submission				
JOURNAL		Submitted (05-MAY-1997) Plant Pathology, The Norwegian Crop				
		Research Institute, Aas 1432, Norway				
FEATURES		Location/Qualifiers				
		1..5873				
		/organism="Lactobacillus sake"				
		/strain="LTH673"				
		/db_xref="taxon:1599"				

```

CDS
complement(1148..1495)
/note="Orfy"
/codon_start=1
/transl_table=11
/protein_id="AAB93966.1"
/db_xref="PID:g2735686"
/db_xref="GI:2735686"
/translation="MTNQEIAKENITHLYNSLAHPDKSNALLDITDVLQSVYLTLET
AKNPELVNRLANTYISVGVGFIHLNKSEQLLIDLGAYGORAGWNGVIRGDCITSRAE
FFNYSDARKYARV"
1672..1882
/gene="sppP"
/gene="sppP"
1672..1785
/gene="sppP"
/note="sppP"
/codon_start=1
/transl_table=11
/product="inducing peptide preprotein"
/protein_id="AAB93967.1"
/db_xref="PID:g2735687"
/db_xref="GI:2735687"
/translation="MMIFKKLSEKELQKINGMAGNSSNFHKKIKOIFTHR"
1804..3150
/gene="sppK"
1804..3150
/gene="sppK"
/note="sppK"
/codon_start=1
/transl_table=11
/product="histidine protein kinase homolog"
/protein_id="AAB93968.1"
/db_xref="PID:g2735688"
/db_xref="GI:2735688"
/translation="MLYDVSVMQNFVAILLIFLLYRIQRIKIFKRIILLDILAI
IFSILYIFSDASLVNMLRWLGWHEHOKENIKTTDTANLILIIIVIOILLVAVGTI
ISOTFIKISDFSONLNNSTADITLGLFFFAVLEDFGLFKLNKRRELQHLNQEI
IESLEKQYFIFLFIIVIEILAVNGVGVATILLIIFVLCIGMTFWQMLF
LKAYSIOEADQLVRNQQLDYVNEIOQYTELRFKHDYQNILSFSFAEKGDQQ
QFRAIYELQAPDIOEIOGAVIAQDYLKNDPIRGLYQFRAKQAGVLFKEMT
EPIELATNLLYIRIIGILLDNAIEQAVQETDQLVSCAFLOSDGLIETIENTASOV
KNLQAFSELGYSTKGAGRGTLANVQDLIAKQTNLFLEQIENRKLRLQTLITEET"
3152..3898
/gene="sppR"
3152..3898
/gene="sppR"
/note="sppR"
/codon_start=1
/transl_table=11
/product="response regulator homolog"
/protein_id="AAB93969.1"
/db_xref="PID:g2735689"
/db_xref="GI:2735689"
/translation="MYPVYLLEDLQOQAYIOQIIANTIMINEFAMTLTCAASDTETL
LAAIKDQQLGLFLDEIEDNRQAGLEVATKROMPPFAOIVFITTHEELTLTLERK
IAPDLVLDQDQWAEIKRQILDLIDLLAEKQNEAAAYHRENLFYKIGRPFSLPKFV
VLYLTEKNGPHNLLAVTRKVTFCNGLNAEQYPMFLFCRDKSLVLSNLNIAIDSK
TRSLKFVDSSEAKVSFRKSRRELVAKLQNM"
4045..4230
/gene="sppA"
4045..4230
/gene="sppA"
/note="sppA; bacteriocin"
/codon_start=1
/transl_table=11
/product="sakacin P preprotein"
/protein_id="AAB93970.1"
/db_xref="PID:g2735690"
/db_xref="GI:2735690"
/translation="NEKEFELSKEVTAITGKYGVNGVHCKHSCVTVDWGTAIGNIG
NNAAAWATGGAGWKNK"
4324..4620
/gene="sppA"
4324..4620

```

```

/gene="sppA"
/note="sppA"
/codon_start=1
/transl_table=11
/product="putative sakacin P immunity protein"
/protein_id="AAB93971.1"
/db_xref="PID:g2735691"
/db_xref="GI:2735691"
/translation="MKILWYSGGKORGERANDIIGQLLLDNLHDPKNHEALINLY
ONEIKRKESVPPFILSRMNISTANTIRDRLLITDFQEDKLLLTALSNIYGY"
4971..5873
/gene="sppT"
4971..5873
/gene="sppT"
/note="sppT"
/codon_start=1
/transl_table=11
/product="putative ATP-dependent translocator protein"
/protein_id="AAB93972.1"
/db_xref="PID:g2735692"
/db_xref="GI:2735692"
/translation="MIDRKVIYQVDEMDCVRALAMILKDYSSASLAYLRNEAKTD
LEGFTALGLVTAENLGFETKAIADMSLFEVADLPFPFAHLVKWRATILCLGTRD
KRRCHIIADPDPTVGIAMTOSQFKSEWTGVAFIAPKSPYQPVKQTKASLSEFPT
LMQOKRLVINITLAALLITISILGSYFQIVDTYIPNNMQSTLAIVAIGLIIYGF
QSIFTYASKFFIGGRTAVIN"
BASE COUNT 1930 a 904 c 1174 g 1865 t
ORIGIN

Query Match 100.0%; Score 114; DB 17; Length 5873;
Best Local Similarity 100.0%; Pred. No. 7,64e-44;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1672 ATGATGATATTTAAAAAAGTTCAGAAAAGAAATTCGAAAAATAAACGGTGGTATGCA 1731
QY 1 ATGATGATATTTAAAAAAGTTCAGAAAAGAAATTCGAAAAATAAACGGTGGTATGCA 60
|||||
Db 1732 GGAATAGTTCATATTTATTCATAGATTAACAAATTTTACCCATCGTTAA 1785
QY 61 GGAATAGTTCATATTTATTCATAGATTAACAAATTTTACCCATCGTTAA 114
|||||

RESULT 2
LOCUS LSSAKACNP 7597 bp DNA BCT 18-DEC-1996
DEFINITION L.sake spIA, sppA, sppE, sppK, sppR and sppT genes.
ACCESSION Z48542
NID G695615
VERSION Z48542.1 GI:695615
KEYWORDS sakacin P; spIA gene; sppA gene; sppE gene; sppK gene; sppR gene;
SOURCE Lactobacillus sakei.
ORGANISM Lactobacillus sakei.
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
Lactobacillus.
REFERENCE 1 (bases 1 to 7597)
AUTHORS Huhne,K., Axelsson,L., Holck,A. and Krockel,L.
TITLE Analysis of the sakacin P gene cluster from Lactobacillus sake
JOURNAL Lb674 and its expression in sakacin-negative Lb. sake strains
MEDLINE Microbiology 142 (Pt 6), 1437-1448 (1996)
REFERENCE 2 (bases 1 to 7597)
AUTHORS Holck,A.L., Axelsson,L., Huhne,K. and Krockel,L.
TITLE Purification and cloning of sakacin 674, a bacteriocin from
JOURNAL Lactobacillus sake Lb674
MEDLINE FEMS Microbiol. Lett. 115 (2-3), 143-149 (1994)
REFERENCE 3 (bases 1 to 7597)
AUTHORS Tichaczek,P.S., Vogel,R.F. and Hammes,W.P.
TITLE Cloning and sequencing of sakp encoding sakacin P, the bacteriocin
JOURNAL produced by Lactobacillus sake LTH 673
MEDLINE Microbiology 140 (Pt 2), 361-367 (1994)
94236240

```

```

REMARK (sites)
REFERENCE 4 (bases 1 to 7597)
AUTHORS Huehne, K.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1995) Huehne K., Federal Centre for Meat
Research, Microbiology and Toxicology, E.-C.-Baumann-Strasse 20,
Kulmbach, Bavaria, FRG, D-95326
FEATURES
    source
        1..7597
            /organism="Lactobacillus sakei"
            /strain="Lb674"
            /db_xref="taxon:1599"
            412..417
            /db_xref="taxon:1599"
            439..444
            /db_xref="taxon:1599"
            557..562
            /db_xref="taxon:1599"
            572..1918
            /db_xref="taxon:1599"
            572..1918
            /db_xref="taxon:1599"
            /gene="sppk"
            /function="putative histidine protein kinase"
            /citation=[1]
            /codon_start=1
            /transl_table=11
            /product="SppK"
            /protein_id="CAA88426.1"
            /db_xref="PID:g695616"
            /db_xref="GI:695616"
            /db_xref="SPTREMBL:Q48870"
            /translation="MLVTVSVSLQMNFVAILLFLLRYIQRKIKFKRIILDLIAI
            IFSILYIFISDASLLVWMLRWLGHQKQENKIKTDTANGLIILVQLLVAVGTI
            ISQTSIIKSFISQNLNNSATDITLLGFFAVFDGLFFILLKRNKRWELQHLNQEI
            IEFLEQYIFIFILFVIEIILAVNLQGVATITLLTIIIFCVLIGMTFWQVNL
            LKAYIQEANDOLVRNQOQDVNIEQOYTTLRRFKHDYQNIILSLGFAEKQDQO
            QKRAYIQEANDOLVRNQOQDVNIEQOYTTLRRFKHDYQNIILSLGFAEKQDQO
            EPIELANLLTVIRIIGILLDNAEQAVETQLYSCAFLOQSDGLIETIENTASQV
            KNLQAFSELGYSKAGRGIGLANVQDLIAQNLFLQYENKRLKRLQILMITEET"
            1906..1910
            /gene="sppk"
            1920..2666
            /gene="sppk"
            1920..2666
            /gene="sppk"
            /gene="sppk"
            /function="putative response regulator protein"
            /note="ttg start codon"
            /citation=[1]
            /codon_start=1
            /transl_table=11
            /product="SppR"
            /protein_id="CAA88427.1"
            /db_xref="PID:g695617"
            /db_xref="GI:695617"
            /db_xref="SPTREMBL:Q48871"
            /translation="MYPVVLEDDLQQAIVQIILANTIMINEFAMTLCAASDTETL
            LAKIQOQUGLFUMDIEDNRQGLVATIKQMPFAQIVFITTHEELTLTLERK
            IAPLYLKKQOTAEIKRQLTDLLAEKQNEAAVHRENLFSYKIGPRFFSPLEK
            VYLYTERNGCHINLLATRVKTPVPGNLNAEQYPMLEFRCDKSYLVNLSNIANYDK
            TRSLKFDGSGAKVSFRKSLVAKLQMM"
            2670..2706
            /terminator
            2712..2717
            /-35_signal
            2736..2741
            /-10_signal
            2798..2802
            /RBS
            2813..2998
            /gene="sppa"
            2813..2998
            /gene="sppa"
            /gene="sppa"
            /standard_name="sakacin P preprotein"
            /function="bacteriocin preprotein"
            /citation=[1]
            /citation=[2]
            /citation=[3]
            /codon_start=1
            /transl_table=11
            /product="SppA"

sig_peptide
    2813..2866
        /gene="sppa"
        /function="signal peptide"
        2867..2995
        /gene="sppa"
        /function="bacteriocin"
        /evidence=experimental
        /product="sakacin P"
        3079..3084
        /gene="sppa"
        3092..3388
        /gene="sppa"
        3092..3388
        /gene="sppa"
        /function="putative protein conferring immunity to sakacin
        P"
        /citation=[1]
        /citation=[3]
        /codon_start=1
        /transl_table=11
        /product="SpiA"
        /protein_id="CAA88429.1"
        /db_xref="PID:g695619"
        /db_xref="GI:695619"
        /db_xref="SPTREMBL:Q57161"
        /translation="MKILKWSGKGRANDIIGQLLDLHDPKNEHLEAILNY
        ONEIKRKSSVPFILLRMINSTANTIRDRLLITDFQEDKLLTALSNIQY"
        3544..3594
        /terminator
        3654..3659
        /-35_signal
        3675..3680
        /-10_signal
        3728..3732
        /RBS
        3739..5895
        /gene="sppT"
        3739..5895
        /gene="sppT"
        /gene="sppT"
        /function="putative ATP-dependent translocator protein"
        /citation=[1]
        /codon_start=1
        /transl_table=11
        /product="SppT"
        /protein_id="CAA88430.1"
        /db_xref="PID:g695620"
        /db_xref="GI:695620"
        /db_xref="SPTREMBL:Q48872"
        /translation="MIDRKIYIPQVDEMDCGVAALMILKHYSASLALRNEAKTD
        LEGTALGLVTAENLGPETKAIQADMSLFEVADLPFPFIAHLKNGELLHYVVLGI
        KDAIILADPDPTVGIKMTOSQKSEWTVALTAPKPSQPVKQPKASLFSFVPT
        LMQKRLVNTLAALLTIISLGSYFLQVVDITYIPNNQMSLAIVAIGLIIYGF
        QSIFYAQNFLLAVLQRLSIEILSYLRHVFLPMSFATRKTEIISRFENDANKII
        DALASAIISLIDITVVVIMGIIITLQNTLFWLTAVIPIYIAVITLFAKPFKLNQ
        KEMESNVNDAIIEDLHGIIKALTSESYQIDAFVDYLLKSFYVKAQTLQO
        ATKLVQLGNVFLVWCAILVTHNELSIGOLMTYNALLAVFNPLQNIINLOTKLOS
        AFVANNVLYLVDSEKTERPKSESQVLPYELQNVYQYGVQGVRLSISLAKIE
        PNEKTIYVMGSGSKSTLVKLLVNFEPTEKLFNQDQITDKTKLRQVNYIAQT
        PYIFSGIWDNLKLNRPGITQADIMHACELAKNDIEKMPLOPETLDDGNTLSG
        GOKRLTARALLSPARVLIIFDESTGLDTEGLLVLDHMLALEERTIIFIAHLAIA
        KKTNNILVLENGSVVETGTGHDALVDVTQGRYQLLNH"
        5897..5901
        /RBS
        5910..7286
        /gene="sppE"
        5910..7286
        /gene="sppE"
        /function="putative accessory protein for sakacin P
        export"
        /citation=[1]
        /codon_start=1
        /transl_table=11
        /product="SppE"

```

[illegible]

RESULT	LOCUS	DEFINITION	DNA	PRI	24-JUL-1998
	AC004787	1	HUAC004787 216021 bp		
	AC004787	2	Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete sequence.		
	AC004787.1	3	GI:3337381		
	HTG	4			
	human.	5			
	ORGANISM	6	Homo sapiens		
	REFERENCE	7	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	AUTHORS	8	1 (bases 1 to 216021)		
	TITLE	9	Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J., Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.		
	JOURNAL	10	Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10		
	REFERENCE	11	Unpublished		
	AUTHORS	12	2 (bases 1 to 216021)		
	TITLE	13	Adams,M.D. and Loftus,B.J.		
	JOURNAL	14	Direct Submission		
	COMMENT	15	Submitted (24-JUL-1998) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA		
		16	On Jul 24, 1998 this sequence version replaced gi:3241936.		
		17	Address all correspondence to: Mark Adams The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA e-mail address: humen@tigr.org. The orientation of the sequence is from SP6 end to 17 end. Genes were identified by a combination of five methods including: XGRAIL (available by an anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/~chris/GENSCANW.html) searches of the complete sequence against a peptide database, and the Human Gene Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).		
		18	Genes without peptide homology having significant EST hits are termed 'Unknown gene product'. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).		
	FEATURES	19	Location/Qualifiers		
	source	20	1. .216021		
		21	/organism="Homo sapiens"		
		22	/db_xref="taxon:9606"		
		23	/chromosome="16"		
		24	/map="#16q21-22"		
		25	/clone="A-952F10"		
		26	27765. .27872		
	STS	27	/notes="7766, STS1-CSRL-27g3-uA/CSRL-27g3-uZ, Chr. -, Homo sapiens"		
		28	/db_xref="dbSTS:G02280"		
	STS	29	73826. .73943		
		30	/notes="7608, STS1-CSRL-24g1-uA/CSRL-24g1-uZ, Chr. -, Homo sapiens"		
		31	/db_xref="dbSTS:G02122"		
	STS	32	175801. 175945		
		33	/notes="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"		
		34	/db_xref="dbSTS:G09703"		
	STS	35	175810. 175945		
		36	/notes="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"		
		37	/db_xref="dbSTS:G09935"		
	STS	38	19463. 195952		
		39	/notes="9824, WI-3555, Chr. 16, Homo sapiens"		
		40	/db_xref="dbSTS:G04338"		
	BASE COUNT	41	60960 a 51778 c 49172 g 53987 t 124 others		
	ORIGIN	42			
		43			
		44			
		45			
		46			
		47			
		48			
		49			
		50			
		51			
		52			
		53			
		54			
		55			
		56			
		57			
		58			
		59			
		60			
		61			
		62			
		63			
		64			
		65			
		66			
		67			
		68			
		69			
		70			
		71			
		72			
		73			
		74			
		75			
		76			
		77			
		78			
		79			
		80			
		81			
		82			
		83			
		84			
		85			
		86			
		87			
		88			
		89			
		90			
		91			
		92			
		93			
		9			

Db	1482	Y Y C Y Y Y Y Y Y Y C M G A M A A A A N Y Y K R S C H A N Y Y Y M G R A R K Y Y Y M A R R R G C A R S Y K	1541
Cp	97	T T T G T T T A A C T T A T G A A T A A A A A T A G A A C T T T T C C T G C A T A C C A C C G T T A T T T T T	38
Db	1542	Y M A Y W A A A T T W C A A A A A A N Y T T Y Y W M M W Y K K W Y	1578
Cp	37	G C A A T C T T T T T C T G A A A G T T T T T T A A A T A T A T C A T C A T	1

6	RESULT	I66494	7218 bp	DNA	PAT	23-DEC-1997
	LOCUS	Sequence 14	from patent US 5670367.			
	DEFINITION	I66494				
	ACCESSION	I66494				
	NID	92724471				
	VERSION	I66494.1	GI:2724471			
	KEYWORDS					
	SOURCE	Unknown.				
	ORGANISM	Unknown.				
	REFERENCE	Unclassified.				
	AUTHORS	1 (bases 1 to 7218)				
	TITLE	Dorner F., Schefflinger, F. and Falkner, F. Gunter.				
	JOURNAL	Recombinant fowlpox virus				
	FEATURES	Patent: US 5670367-A 14 23-SEP-1997;				
		Location/Qualifiers				
		1..7218				
	source	/organism="unknown"				
	BASE COUNT	1944 a	1491 c	1486 g	1929 t	
	ORIGIN				368 others	

```

Query Match      23.7%  Score 27  DB 25: Length 7218;
Best Local Similarity 3.4%  Pred. No. 1.76e+00;
Matches         2;  Conservative 41; Mismatches 16; Indels 0; Gaps 0;

Db 1064  GATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1122
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 71  GAACATATTCCTGCATATACACACGCTATTTATTTTGGAAATCTTTTTCTGAAAGTTTTT 13

```

RESULT	7	LOCUS	MVU87256	1056 bp	DNA	MAM	02-JAN-1999
DEFINITION		Mustela vison	GT dinucleotide repeat,				chromosome 1q.
ACCESSION		U87256					
NID		94099442					
VERSION		U87256.1	GI:4099442				
KEYWORDS		American mink.					
SOURCE		Mustela vison					
ORGANISM		Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;					
		Carnivora; Fissipedia; Mustelidae; Mustela.					
REFERENCE		1 (bases 1 to 1056)					
AUTHORS		Brusegaard, K., Shukri, N.M., Malchenko, S., Koroleva, I. and Lohli, O.					
TITLE		Direct Submission					
JOURNAL		Submitted (27-JAN-1997) Breeding and Genetics, Danish Institute of					
		Animal Science, Blichersalle K25, Tjele 8830, DK					
FEATURES		Location/Qualifiers					
SOURCE		1..1056					

```

/map="1q"
/note="primers: 1167F: agccctgcatactactctt, 1167R:
gagatcttaccgctgttgag"
98 119
/standard_name="1167F"
primer_bind complement(300..320)
/standard_name="1167R"
BASE COUNT 211 a 221 c 210 g 225 t 189 others
ORIGIN

Query Match 22.8% Score 26; DB 23; Length 1056;
Best Local Similarity 15.9%; Pred. No. 4.43e+00;
Matches 18; Conservative 59; Mismatches 33; Indels 3; Gaps 3;

```

[illegible]

RESULT	8	CAU072146	7430 bp	DNA	PLN	09-SEP-1997
LOCUS		Camptotheca acuminata	3-hydroxy-3-methylglutaryl coenzyme A			
DEFINITION		reductase gene, complete cds.				
ACCESSION		U72146				
NID		g1763233				
VERSION		U72146.1	GI:1763233			
KEYWORDS						
SOURCE		Camptotheca acuminata.				
ORGANISM		Camptotheca acuminata.				
		Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;				
		Embryophyta; vascular plants; seed plants; Magnoliophyta;				
		Magnoliopsida; Cornaceae; Cornales; Cornaceae; Cornaceae;				
		Nyssoidae; Camptotheca.				
REFERENCE		1 (bases 1 to 7430)				
AUTHORS		Maldonado-Mendoza,I.E., Vincent,R.M. and Nessler,C.L.				
TITLE		Molecular characterization of three differentially expressed				
		members of the Camptotheca acuminata 3-hydroxy-3-methylglutaryl				
		reductase (HMGR) gene family				
		Plant Mol. Biol. 34 (5), 781-790 (1997)				
JOURNAL		97422400				
MEDLINE		2 (bases 1 to 7430)				
REFERENCE		Maldonado-Mendoza,I.E. and Nessler,C.L.				
AUTHORS		Direct Submission				
TITLE		Submitted (21-SEP-1996) Biology, Texas A & M University, College				
JOURNAL		Station, TX 77843-3258, USA				
FEATURES		Location/Qualifiers				
		1..7430				
		source				

```

/EC_number=1.1.1.34"
/notes="converts HMGCoA in mevalonate precursor for
isoprenoid compounds; HMGR; hmg2; similar to protein
encoded by GenBank Accession Number L10390"
/codon_start=1
/product="3-hydroxy-3-methylglutaryl coenzyme A reductase"
/protein_id="AA569727.1"
/db_xref="PID:g1763234"
/db_xref="GI:11763234"
/translation="MDVRRRPTKSLRPKAKTAHAAGEPLKPHQHNSLSKASDALPLPLY
LTNGVFETFTFSVAYLHWRREKIRKSTPLHIVTVELAALISLTSVYILGFEFGI
DFQSIITRASHDSVEVDNDQRFLEEDSRPGCAAADRLSLVPPQAAVAAPKPK
VVEDPVPVLTATSEEDDEILKVVAGTTPSYLSRSLGNCKRAAIIPOVALAQLTKGS
LAGPLDGEDYESILGOCCEMAYGVQMAVGIAGPLLLDGRVYLVPMAITFEGCLVAST
NRCKAIYASGGATSVILRDQWTPRAPVPGTKAKSAELKEFLEDPNFTDLAVFNKK
SRFRPFGQTHCALQNLFMRESCSTDGAGNMWMSKGVQNVLDLFQNDPDSMDVILG
ISGNFCSDRKPAAVNIETGRKSVGCEIIEEEVVRKVLTVNASVLKLNKLNLTGS
AMAGALQNVESHASNTSVAVYIATGDQPAQNVESHCHITMMEAVNDGDKDLHVSVSMPS
EYIVGVGGTQLLQSGACNLGLVKGSKESPGSNRLLATIVAGSVLAGELSLMSAIA
AGOLVKSHMKYNSRCSDDITKVS"
2300 a 1276 c 1377 g 2477 t

```

```

Query Match      22.8%   Score 26;   DB 28;   Length 7430;
Best Local Similarity 82.5%;   Pred. No. 4.43e+00;
Matches 33;   Conservative 0;   Mismatches 7;   Indels 0;   Gaps 0;

Db 1433 TAAAAATATATTTTAAATATTTATTTATTTAAAAATTTAAAAATATT 1472
Cp 103 TAAAAATTTCTTTTAACTCTTTATGAATTTAAAAATTTAGAACTATT 64

```

```

RESULT 9
LOCUS AF012089 10772 bp DNA INV 05-AUG-1997
Drosophila melanogaster cysteine proteinase-1 (Cpl) gene, complete
cds, and phenylalanine tRNA synthetase gene, partial cds.
ACCESSION AF012089
NID 92305220
VERSION AF012089.1 GI:2305220
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Eukaryota; Insecta; Pterygota; Diptera; Brachycera;
Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 4546 to 4553)
Gray, Y.H., Tanaka, M.M. and Sved, J.A.
P-element-induced recombination in Drosophila melanogaster: hybrid
element insertion
Genetics 144 (4), 1601-1610 (1996)
77132596
2 (bases 1 to 10772)
Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
Structure of the cysteine proteinase (Cpl) gene of Drosophila
melanogaster and associated mutational effects
Unpublished
3 (bases 1 to 10772)
Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
Direct Submission
Submitted (30-JUN-1997) School of Biological Sciences, University
of Sydney, Biology A12, Sydney University, NSW 2006, Australia
Location/Qualifiers
1..10772
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
join(872..1000,2310..2426,6476..6690,6751..7707)
/gene="Cpl"
/product="cysteine protease"
872..7707
/gene="Cpl"
872..1000
/gene="Cpl"
/number=1
1001..2309
/gene="Cpl"
2310..2426
/gene="Cpl"
/number=2
join(2328..2426,6476..6690,6751..7462)
/gene="Cpl"
/codon_start=1
/product="cysteine proteinase-1"
/protein_id="AAB65749.1"
/db_xref="PID:g2305221"
/db_xref="GI:2305221"
/translaton="MRTAVLLPALLAVAAQVSPADVVVNEWHFKLEHKNYQDET
EERFLKFNENKHTAKHNORFAEGKVSFKLVAKNYADLLHFERQLMNGFNITLHK
OLRADESFGVTFISPAHVLPKSDVWRTKGAATVAKDQGHGSCWAFSSGALLEGQ
HFRSGVLSQNLVDCSTKYGNNGCNGLMDNAFRYKDNIGDITKESYFEAD
DSCFNKGTGATDRTDIPQDQEKMAEAVATVGPVSAVDASGHESFQYSEGVN
EPQDAQNLHDHGLVGVFGTDESGDYLVKNVSGTGTGDKGFIKMLRNKNQCGIAS
ASSYPLV"
2427..6475
/gene="Cpl"
4546..4553
/gene="Cpl"
/note="insertion site of P[Casper]{50C}."
/citation={1}
6476..6690
/gene="Cpl"
/number=3
6691..6750
/gene="Cpl"
6751..7707
/gene="Cpl"

```

```

/number=4
join(<8110..9300,9370..>9532)
/product="phenylalanine tRNA synthetase"
join(8110..9300,9370..>9532)
/note="potential orf"
/codon_start=1
/product="phenylalanine tRNA synthetase"
/protein_id="AAB65750.1"
/db_xref="PID:g2305222"
/db_xref="GI:2305222"
/translaton="MLLTLRQGARHLKSTRCLASSAAPKSPSPQLEVSSTVYA
TDGNTVNPKILSYVGAANKHLQTDHPLSIIRQIVNVFYGAYRNQGNPLFSVDOMN
PVYTVQNFQDNLIPADHVSQKSCYVYINQHLRLRAHTAHQVELISGGDLNVLVG
EYVRDEIDSTHYFVHQADAVRLVTKDLFERNPGLFEETWSGLADPKLLPHP
SWTAPNLPATRAKVMHEKMKHLVGLTKDLFGPRIKRYMVDTPFTQPSWELEI
YFKDNMLVGLCGIMRHEILQIRSGVHQISGYAFGVGLERLAWLFDIDIRLFSNDS
GFLSQFSEKDLNLPKYPISHPYQCTNDLSFWLQDIEVDAGFSPNDFYDLVRSVAG
DMVEQISLVDFKPKTKCKSSVCFRIVYRHMTLTQAEVNEIHKQIASASVDSENVQ
IX"
BASE COUNT 2929 a 2357 c 2282 g 3046 t 158 others
ORIGIN
Query Match 22.8%; Score 26; DB 21; Length 10772;
Best Local Similarity 19.7%; Pred. No. 4.43e+00;
Matches 13; Conservative 33; Mismatches 20; Indels 0; Gaps 0;
Db 1693 RWYWAATTTMKWMTTWKWKMMKTYRTWKKWYWSRTTTTSSAMMMWYTWSTWYK 1752
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 71 GAACATATTCCTGCCATACACCGTTATTTTTCGAATCTTTTCTGAAAGTTTATA 12
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1753 WAYAWM 1758
: : : : :
Cp 11 AATATC 6
RESULT 10
LOCUS CELY50F7A 31933 bp DNA INV 06-FEB-1999
DEFINITION Caenorhabditis elegans cosmid Y50F7A.
ACCESSION AF125457
NID 94226117
VERSION AF125457.1 GI:4226117
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1. (bases 1 to 31933)
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lighthouse, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
Saunders, D., Shownkeen, R., Smalton, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Woldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2. (bases 1 to 31933)
Strong, C., Tin-Wollam, A. and Harper, M.
AUTHORS The sequence of C. elegans cosmid Y50F7A
JOURNAL Unpublished (1999)
REFERENCE 3. (bases 1 to 31933)
Waterston, R.
AUTHORS Direct Submission
JOURNAL Submitted (03-FEB-1999) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center

```



```
/db_xref="PID:g3880954"
/db_xref="GI:3880954"
/translation="MKAVFLVFWALFVGVQVQVDFVMMDEKASNPGRSALIYQRK
SYLOPFYNAPQLQFWRRY"
19222 20371
/gene="Y43F8A.2"
join(16922..17061,17112..17207,17256..17374,18018..18216,
18259..18442,19113..19324,19375..19582,20065..20196,
20276..20371)
/gene="Y43F8A.2"
/codon_start=1
/protein_id="CAA21643.1"
/db_xref="PID:e1359893"
/db_xref="PID:g4008433"
/db_xref="GI:4008433"
/translation="MPTDAPTAAPNPPPPAAQOKKLQYDFSGPLTELOGSYLKS
RQSQARPSWASVSWVETERNIGIFWPNCHAYVLCVLPKALIDGENQHDSDFC
DTNVIKYPHISKLKYPKQGAESYFVKSQLOFQKSDKHVLEALPVGHYEMK
KQEDFKLEIDFEGKDLILLNSRCEVKNKPSRPHFIDFETDKNGTIDKSWPN
WLFKNTQITDPSLPALGREGKSIDLEFLKSTVAGNKLGNIVQKACDKIGITKL
RKPAFERAVSIALSYEIFEANTQANETTREYFVSNCEWIRIMASDKTMKEYATAM
KVDTELQVLSAGVYFPPDSQALPLPDYHLIDKLKLTVMKDPVAGDNHLEIFAFG
SVNAVYLFERHRSKAFGVKNLHVIRSEKCEANTAVKPSFYAGISNPEYINIKQVRS
CEFNYYTEPS"
21629..225159
/gene="Y43F8A.3"
join(21629..21805,22493..22731,23411..23784,23925..24052,
24742..24981,25028..25159)
/gene="Y43F8A.3"
/notes="predicted using GeneFinder"
/codon_start=1
/protein_id="CAA21641.1"
/db_xref="PID:e1350422"
/db_xref="PID:g3880956"
/db_xref="GI:3880956"
/translation="MQPLSKFSITFATILLAFSILLVLHKLPLGLCEKSTVDVWV
IHLPFLRAFYTPANIETTPANQWTRYSGLLAKTMGPVTLFERLVRVTDAT
WNGVHVTVPRVSNVDSGAVFIHGGGFAIGSAVAINQVSNIFFFNFKLVGSMNTFV
VSDIRLSPTEVPENLLDCEKAIDYFLENSKFKIDPKKVLVGDSDAGNLAATA
QRAEKAELPQLVLLVPLLOLVQNTSYRFFHKLRTGAFVDPVPSVAVMYLXA
GIPLEKAKELVTLVNGHVKPDYREKIDKLTYRTISTSTHYNTIKPKWEIVEN
SEAQNLLEKVFIDPNSPINRENLENLPKSLIVTCYDVLDRDEGLIYSERLSKSGVPT
KLINXNGYHAMLMNHNEITEASTCLDDVNMHWLIEQ"
25447..28939
/gene="Y43F8A.4"
join(25447..25745,26752..27030,27678..28025,28756..28939)
/gene="Y43F8A.4"
/notes="predicted using GeneFinder"
/codon_start=1
/protein_id="CAA21642.1"
/db_xref="PID:e1350423"
/db_xref="PID:g3880957"
/db_xref="GI:3880957"
/translation="MLEEYFIEVKPQEQQLNLLQNSQWAVFCSNLQTFMFFFG
VFENLHIFILLKSRQNTNVLIIIGIAISDIFLYIVVEGTRFENGIPGKCRP
KKEFLAYIWIITIFEDVRVFSAGSVSLALRYLVNMGARTINRIVGTCRPP
LFSTFLSVISSINHVRYVMRYPDWPPSCRMYPPTLPWFEQVNDPLDGFQ
RQISTFLYIDGVFKIIPILYPLFLAFGLLWELKARDSRKILMRKGEHEMVHTKL
VIFMTGYFLAETPVGISYFVLAAYNGEDFGIIFLANNITVIFVTFLLINSIHCFEC
FFLSOYRLAFWIFGCKCASSGSSVSNRSMSS"
complement(29231..34299)
/gene="Y43F8A.5"
complement(join(29231..29405,30014..30170,30410..30643,
30712..30895,31865..32031,32228..32390,32663..32813,
33997..34109,34216..34299))
/gene="Y43F8A.5"
/notes="predicted using GeneFinder; cDNA EST Y41694.5
comes from this gene; cDNA EST Y41694.3 comes from this
gene"
/codon_start=1
/protein_id="CAA21644.1"
/db_xref="PID:e1359894"
/db_xref="PID:g4008434"
```

```
/db_xref="GI:4008434"
/translation="MGVITITFLYALTSSVFPVQSLIFKYKACITLNSNTEPETACK
SKETIARONSVMSSVILMASSSTGLCVSAFTSRWGHLSDVKSRLAFLIPFGLF
ISDETILIOVLPRLSPYFVYFVYGGFSGYMSITGAFSIVTMSHDORERAKGI
ARLEGTISLSTGVGLISSQKNGFYGLAAFFLVAHFAFMSAVFMDPKPDLOEK
CVKRPFGCHGTDLLKDKRPPTNLKILYFSAFACVFAFICGTRILFYLKHFPFMA
EEFGYLVKAMNTGMTTMSIALYLPFLKNIGIDIRLAIKGLTRISGRAMYAVNEGWT
VEIVVFPFMSKFPATARRSSIATNVGHERGLAFSLVAEALCNLTSSWVFLHAWP
LSLVNFPQSLFVIMPIILIPAILTMINLNRYLEAPLNPEDLTTSIKKSDDISENVY
HDPEALPVLIHIPTETDTLTDSTA"
37699..37771
/gene="Y43F8A.t1"
/notes="ATG His H-tRNA; predicted using tRNAScan-SE-1.1.1;
preliminary prediction; similar to tRNA-His"
37699..37771
/gene="Y43F8A.t1"
BASE COUNT 13158 a 6527 c 6764 g 12407 t
ORIGIN
Query Match 22.8%; Score 26; DB 21; Length 38856;
Best Local Similarity 86.1%; Pred. No. 4.43e+00;
Matches 31; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 32092 AAAAAATTCGAAAAAAATTTCAAAAAA 32127
||||| ||| ||||| ||||| ||||| |||||
QY 13 AAAAAATTCGAAAAAAGAAATTCGAAAAAATAAC 48
||||| ||| ||||| ||||| ||||| |||||
RESULT 12
LOCUS CEC36B1 44996 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid C36B1, complete sequence.
NID 280215
VERSION g1546718
KEYWORDS 280215.1 GI:1546718
HTG; C3HC4; Di-hydrofolate reductase; DNA directed RNA polymerase
II; Fibronectin type-III domains; Proteasome A-type subunit; Serine
threonine protein kinase; U4/U6 small nuclear ribonucleoprotein;
Zinc finger.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 44996)
Lennard, N.
REFERENCE Direct Submission
AUTHORS Submitted (14-SEP-1996) Louis, MO 63110, USA. E-mail:
jesesanger.ac.uk or rwenematode.wustl.edu
2 (bases 1 to 44996)
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,
Saunders, D., Showkeen, R., Smaiden, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J., and Wohldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
COMMENT Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:
http://webc.sanger.ac.uk/cgi-
bin/display?db=wormace&class=Sequence&object=C36B1
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
```

dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 This sequence is the entire insert of clone C36B1. The true right end of clone p39H11 is at 17898 in this sequence. The start of this sequence (.1110) overlaps with the end of sequence 281079. The end of this sequence (44893..44996) overlaps with the start of sequence 236047.

FEATURES

Source

Location/Qualifiers

1..44996
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="I"

gene

/clone="C36B1"

CDS

/complement(506..5930)
 /gene="C36B1.1"
 complement(join(506..619,868..1025,1081..1154,1205..1302,1357..1466,5816..5930))
 /note="CDNA EST EMBL:D34750 comes from this gene"

/codon_start=1

/protein_id="CAB02278.1"

/db_xref="PID:el344394"

/db_xref="PID:g3874785"

/db_xref="GI:3874785"

/db_xref="SPTREMBL:Q93336"

/translation="MLPRHIGGLYFCALYLLRQASQPTFFALNALPGDSETEV
 DILAPKASLNADARVAKGIESPAIGLQGVETAVPYRLYPRNFKQFSLATI
 KLPADKRGVYFAAVNADSDVIGLIEPAGTKQTNISLIVRSVAIVSLFVDFDSQW
 TOFALEVDTVTFFKCRFASRQVSLPDSFDEAKLYIASAGPIIDNGFEVSLF
 LILL"

/complement(7075..10843)

/gene="C36B1.2"

complement(join(7075..7179,7290..7366,7941..8189,
 8242..8411,8742..9000,9134..9328,10759..10843))

/note="similar to fibronectin type-III domain"

/codon_start=1

/protein_id="CAB02267.1"

/db_xref="PID:el344386"

/db_xref="PID:g3874774"

/db_xref="GI:3874774"

/db_xref="SPTREMBL:Q93337"

/translation="MIETGSQLLLFLFTCLSFNSSTALQHEHDVNPAPQNVRIKTO
 STSATLWMDAPPPTVLIRGITVEIGESISQKILLEGPDSTFTVIRLSPNINYYFA
 VSAVNAEGEDGPKVMAAKTRPESQTEKLPPTSVRARIDERSAAGSAFVSWDDP
 NPSSSENSTQKQYVINGYIESDTQOKVRSNAKAVRLTLGIPKEYEYAVKYVA
 GDGRSPWSIRDLVPEKTVSKFDFWCLNDTMCSTHSSPHWKLCKEKKHDTYTOR
 DAGACPRVOYPPSPAHLTTPAINLPDAQRLCLYFALLNHPGQMKVEIFRDGDMAN
 KQIVKTRANVSYSFLKWGQOPRIVIVHEMDVLSGCSERSGK"

/complement(13903..15051)

/gene="C36B1.4"

complement(join(13903..14016,14097..14393,14440..14691,
 14933..15051))

/note="similar to proteasome A-type submit: CDNA EST
 EMBL:M79843 comes from this gene; CDNA EST EMBL:D74896
 comes from this gene; CDNA EST yk361a1.3 comes from this
 gene; CDNA EST yk361a1.5 comes from this gene; CDNA EST
 yk412d8.3 comes from this gene; CDNA EST yk412d8.5 comes
 from this gene"

/codon_start=1

/protein_id="CAB02269.1"

/db_xref="PID:el344385"

/db_xref="PID:g3874776"

/db_xref="GI:3874776"

/db_xref="SPTREMBL:Q95005"

/translation="MRYDRAIIFSPDGHFOVEAQAQVKKGSTAVGVKDCITVI
 GVFKSIPALQDRTIRKIMIDHVMYLAFAGISADARVLDVDRARIEQSKYKLTLED
 VTVAYISRYTANTKQRTSQSPRRPFGISMLIGFDHDTGPRFLRKFEPGAYEYVAN

gene

CDS

ATGRGKPVREYLEOYSEENTVDEATTLKLVKSLAQVPPGSGNQIEIAVMKKVND
 LQORVLSSTEEIALKVVETERRAAEAEASKKK"
 15333..17327
 /gene="C36B1.3"
 join(15333..15455,15510..15644,15692..15817,15970..16217,
 16267..16355,16401..16524,16659..16865,16912..16963,
 17223..17327)
 /gene="C36B1.3"
 /note="similar to DNA directed RNA polymerase II: CDNA EST
 EMBL:T00665 comes from this gene; CDNA EST EMBL:T01799
 comes from this gene; CDNA EST EMBL:T01866 comes from this
 gene"

/codon_start=1

/protein_id="CAB02268.1"

/db_xref="PID:el344384"

/db_xref="PID:g3874775"

/db_xref="GI:3874775"

/db_xref="SPTREMBL:Q93338"

/translation="MPYANQNIETVLTNDIKFVLWDTLSVANSRRVFAEVPY
 LAIDWQIETNTSVLHDEFIAHMGILPISDYHVKMQYTRDCEAEFCDECSYPTI
 LQMKCKDEATLAVTTEHLMPHNLTITVRPACGKALRERGSTRDFEHNEEILIVKLR
 KQELNLKAYVKKGSGEKHAKWNPCTGVAFEDPDNALRHTIYPNVEEPRSDHSLP
 EDSTEKEAFEPDNPKNFWSIEGTGALPAQRIYVMGIGILKRLLELMAISNELQ
 AHAQYHILIFHEVKKHTSYNTAFFSONMVGDCERKLTIVGVDPYRNKKVNRNADGS
 GKPTVTKNRLIGVKKATIVGAKCKLCKMLIHQPGSHYCSCTCAYQKGCAMCGKRIQ
 NTKGLRQSTT"

18732..19246

/gene="C36B1.6"

join(18732..18934,18987..19033,19065..19246)

/gene="C36B1.6"

/codon_start=1

/protein_id="CAB02271.1"

/db_xref="PID:el344387"

/db_xref="PID:g3874778"

/db_xref="GI:3874778"

/db_xref="SPTREMBL:Q93340"

/translation="MPGKECVILFVSVLTVLHNFMMKFNSFEDKITCDTHFRLLHS
 VKTQYLEYFPNNGNYNCLYCTMCLDFSSNCSIDVFNISNIPMIMICIASLMVQI
 VMESFPCKVDRSVKMKMECSIDCCSIFSVLLNIAIRVMVEGG"

complement(20092..22271)

/gene="C36B1.5"

complement(join(20092..20199,20247..20454,20533..20763,
 21060..21177,21227..21394,21441..21720,21763..21900,
 21967..22119,22185..22271))

/note="similar to U4/U6 small nuclear ribonucleoprotein"

/codon_start=1

/protein_id="CAB02270.1"

/db_xref="PID:el344386"

/db_xref="PID:g3874777"

/db_xref="GI:3874777"

/db_xref="SPTREMBL:Q93339"

/translation="NAENGNEAVPAPPROFGSLANAEVNAIILNAOQNHGPTVSLER
 MEVSNQADSRHDAEMFAEDRRRARTLTPTDDVQVQLKRLALNOPICLFGDADLR
 RKLRLALLSTDEIAVAHITDEVNADKADSETVYHGRGPIELRMARVSIADFSIR
 RKLRLDKAREEAPAAHAKALQAQHVQVQIINHSAQVADRVAFVCEFSADSEH
 IYTAGWSQVAVKKEQCAQEKFTGHSSQAGCARFPGAFQNDYSLSNVSCSDYQ
 TVLLWSLSQESPIGEQHPQVRSKVAHPNGHHLATACFDSTWRMYDLTTKRELLYQ
 EKHGKSADVAFAHPDGSVALTGCHDCYGRVMDWMTGRCIMFLDGHKTHSVEMPNQ
 YEMITGSSDMSMKVMDLRNRTYTMPARTSVYTVRADAAGQYLVASASDCTLKMS
 TTGWQPLRLQLQSHDRILCLVDISPDQWMCSSAFDRTFKLWQSDY"

22786..23487

/gene="C36B1.7"

join(22786..22920,22965..23095,23142..23294,23337..23487)

/note="similar to di-hydrofolate reductase: CDNA EST
 EMBL:D72539 comes from this gene; CDNA EST EMBL:D75436
 comes from this gene"

/codon_start=1

/protein_id="CAB02272.1"

/db_xref="PID:el344388"

/db_xref="PID:g3874779"

/db_xref="GI:3874779"

gene

CDS

```

/db_xref="SWISS-PROT:Q93341"
/translation="MRKNLIVAMDAGGIGKNGVLPWRKKDMQYFASVTKNVSOS
KNAVLMGRKQWESIVTRRPLRPLAGRLNIVLSQLPKQSDYIVVNSLEAMKLLSEP
PVDSEIWINIGGAEIYDLALRENLYDEIHLTRFKNFADVHLKSLDFSKMEKVN
AEVSSENSEIPEENGLKFECKWKVVENH"
23636..28511
/gene="C36B1.8"
Join(23636..23782,23832..24032,24475..24625,24685..24872,
24971..25047,25100..25193,25241..25570,25668..25987,
26036..26268,26315..26553,26602..26710,26756..26992,
...
Note: remainder of annotations omitted.

```

```

Query Match      22.8%; Score 26; DB 21; Length 44996;
Best Local Similarity 88.2%; Pred. No. 4.43e+00;
Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 36007 TTTATTCGAATTTTTCGAGAAAGTTTTTAAA 36040
Cp 43 TTTTTCGAATCTTTTCTCGAAAGTTTTTAAA 10

```

```

RESULT 13
LOCUS AC002489 91638 bp DNA HTG 20-JAN-1998
DEFINITION Homo sapiens chromosome X, clone 592, WORKING DRAFT SEQUENCE, 8
unordered pieces.
ACCESSION AC002489
NID 92795820
VERSION AC002489.1 GI:2795820
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 91638)
AUTHORS Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P. and
Lander,B.

```

```

TITLE Homo sapiens chromosome X, clone 592
JOURNAL Unpublished
AUTHORS Hawkins,T.L., Birren,B.W., Fasman,K.H., Nusbaum,C., Lander,E.S.,
McKernan,K., Munro,C., Richardson,P., Barna,N., Chang,A., Cooke,P.,
Daly,M.J., Dewar,K., Forrest,C., Gage,D., Gerlitzky,K.,
Hagos,B., Huang,J., Hui,L., Jacotot,L., Kirby,A., Lane,M.,
Mackenzie,J., Marquis,N., McDermott,J., Molla,M., Morrow,J.,
Nachman,A., Naylor,J., Nusbaum,C., O'Connor,T., Olotu,A.,
Peterson,K., Reeve,M.P., Roberts,D., Rollins,G., Stilwell,J.,
Stone,C., Strickland,C., Sydney,K., Tang,L., Wilmer,F., Zemtseva,I.
and Zody,M.

```

```

Direct Submission
Submitted (22-AUG-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 20, 1998 this sequence version replaced gi:2772532.
The Staden databases, finishing information, and all
chromatographic files used in the assembly of this clone are
available from our anonymous ftp site.

```

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 8591: contig of 8591 bp in length
* gap of unknown length
* 8592 28835: contig of 20244 bp in length
* gap of unknown length
* 28836 40356: contig of 11521 bp in length

```

```

* 40357 58902: contig of 18546 bp in length
* gap of unknown length
* 58903 61812: contig of 2910 bp in length
* gap of unknown length
* 61813 71640: contig of 9828 bp in length
* gap of unknown length
* 71641 75199: contig of 3559 bp in length
* gap of unknown length
* 75200 91638: contig of 16439 bp in length.
FEATURES
Location/Qualifiers
1..91638
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="592"
/clone_lib="unknown"
/map="X"
/chromosome="X"

```

```

BASE COUNT 25406 a 19091 c 18992 g 28011 t 138 others
ORIGIN

```

```

Query Match      22.8%; Score 26; DB 19; Length 91638;
Best Local Similarity 88.2%; Pred. No. 4.43e+00;
Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Db 34433 AAAAACTTTCAGAAAAAGCATTTTAAAGAAATAA 34466
|||||
QY 13 AAAAACTTTCAGAAAAAGCATTTTCAAAAAATAA 46
|||||

```

```

RESULT 14
LOCUS AC005094 158078 bp DNA PRI 27-AUG-1998
DEFINITION Homo sapiens BAC clone RG332B22 from 7p15-p21, complete sequence.
ACCESSION AC005094
NID 93478660
VERSION AC005094.1 GI:3478660
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 158078)
AUTHORS Edwards,J. and Ozersky,P.
TITLE The sequence of Homo sapiens BAC clone RG332B22
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 158078)
AUTHORS Waterston,R.H.
TITLE Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

3 (bases 1 to 158078)
Waterston,R.
Direct Submission
Submitted (27-AUG-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 27, 1998 this sequence version replaced gi:3212897.
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:saplens@watson.wustl.edu

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems,

such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone RG332B22 is from the first release of the human BAC library CITB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:18794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelOBAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RG023M10, 200 bp overlap. Actual start of this clone is at base position 1 of RG332B22; actual end is at 21981 of RG023M10.

FEATURES

source	Location/Qualifiers
STS	1..158078 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7p15-p21" /clone_lib="CITB-978SK-B" /clone="RG332B22" complement(4183..4302) /db_xref="GI:454632"
repeat_region	5685..5885 /rpt_family="Alu"
repeat_region	6316..6774 /rpt_family="L1"
repeat_region	6376..7114 /rpt_family="L2"
repeat_region	7454..7659 /rpt_family="MIR"
repeat_region	7747..8012 /rpt_family="MIR"
repeat_region	8815..9124 /rpt_family="Alu"
repeat_region	9384..9440 /rpt_family="Alu"
repeat_region	10844..11019 /rpt_family="MIR"
repeat_region	11795..12088 /rpt_family="L2"
repeat_region	12304..12600 /rpt_family="L1"
repeat_region	12877..13179 /rpt_family="Alu"
repeat_region	14006..14179 /rpt_family="Alu"
repeat_region	14600..14678 /rpt_family="Alu"
repeat_region	15150..15521 /rpt_family="MER81"
repeat_region	15530..15826 /rpt_family="MaLR"
repeat_region	16749..17049 /rpt_family="Alu"
repeat_region	17053..17133 /rpt_family="Alu"
repeat_region	17157..17290 /rpt_family="MaLR"
repeat_region	17294..17591 /rpt_family="L1"
repeat_region	17617..17896 /rpt_family="Alu"
repeat_region	19068..19369 /rpt_family="MaLR"
repeat_region	19842..20012 /rpt_family="Alu"
repeat_region	20131..20434 /rpt_family="Alu"
repeat_region	21967..22200 /rpt_family="Alu"
repeat_region	23876..24007 /rpt_family="Alu"
repeat_region	24222..25020 /rpt_family="MIR"
repeat_region	25430..25805 /rpt_family="MER21_g"
repeat_region	25806..25860 /rpt_family="MaLR"
repeat_region	25991..26070 /rpt_family="MIR"
repeat_region	26083..26295 /rpt_family="MIR"
repeat_region	26317..26548 /rpt_family="MER1_type"
repeat_region	27048..27148 /rpt_family="MIR"
repeat_region	27169..27432 /rpt_family="Alu"
repeat_region	27585..27790 /rpt_family="Alu"
repeat_region	27852..28204 /rpt_family="L2"
repeat_region	28407..28808 /rpt_family="MER2_type"
repeat_region	29146..29736 /rpt_family="L1"
repeat_region	30340..31208 /rpt_family="MER2_type"
repeat_region	32747..32816 /rpt_family="MER2_type"
repeat_region	33065..33407 /rpt_family="L2"
repeat_region	33470..33529 /rpt_family="L2"
repeat_region	34582..35314 /rpt_family="L1"
repeat_region	35993..36158 /rpt_family="L1"
repeat_region	36171..36218 /rpt_family="Alu"
repeat_region	36219..36392 /rpt_family="L1"
repeat_region	36393..37446 /rpt_family="Alu"
repeat_region	37449..37744 /rpt_family="L1"
repeat_region	37745..37983 /rpt_family="Alu"
repeat_region	38043..38104 /rpt_family="L1"
repeat_region	38554..38728 /rpt_family="L2"
repeat_region	38733..39032 /rpt_family="MER1_type"
repeat_region	39339..39424 /rpt_family="Alu"
repeat_region	39426..39706 /rpt_family="BC200"
repeat_region	40645..40760 /rpt_family="Alu"

MISSREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Wed Sep 15 10:53:11 1999; MasPar time 51.25 Seconds
Tabular output not generated. 476.662 Million cell updates/sec

Title: >US-09-068-507A-2
Description: (1-114) from US09068507A.seq
Perfect Score: 114
N.A. Sequence: 1 ATGATGATATTTAAACT.....AAATTTTACCCATCGTTAA 114
Comp: TACTACTATTAATTTTGA.....TTTAAATGGTGCAATT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 7.049; Variance 6.074; scale 1.161
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description Pred. No.
1	114	100.0	114 34 767141 IF gene for bacteriocin 4.71e-41
2	30	26.3	91 9 051746 Oligonucleotide probe 8.34e-03
3	30	26.3	91 46 V44650 Mammalian DNA replica 8.34e-03
c	30	26.3	91 9 051746 Oligonucleotide probe 8.34e-03
5	29	25.4	204 1 N81164 Base substituted E.co 2.10e-02
c	29	25.4	204 1 N81164 Base substituted E.co 2.10e-02
7	28	24.6	501 3 N50034 Sequence encoding new 5.26e-01
8	27	23.7	91 46 V44650 Mammalian DNA replica 1.31e-01
9	26	22.8	501 3 N50023 Sequence encoding new 3.21e-01

c	10	26	22.8	501 3	N50026	Sequence encoding new	3.21e-01
	11	24	21.1	114 12	Q70467	Generic DNA sequence	1.89e+00
	12	24	21.1	498 3	N50034	Sequence encoding new	1.89e+00
	13	24	21.1	501 3	N50032	Sequence encoding new	1.89e+00
	14	24	21.1	501 3	N50031	Sequence encoding new	1.89e+00
	15	24	21.1	501 3	N50033	Sequence encoding new	1.89e+00
	16	24	21.1	501 3	N50029	Sequence encoding new	1.89e+00
	17	24	21.1	501 3	N50028	Sequence encoding new	1.89e+00
c	18	24	21.1	501 3	N50033	Sequence encoding new	1.89e+00
	19	23	20.2	67 24	T14322	Primer used in the la	4.50e+00
c	20	23	20.2	114 12	Q70465	Generic DNA sequence	4.50e+00
c	21	23	20.2	114 12	Q70469	Generic DNA sequence	4.50e+00
c	22	23	20.2	114 12	Q70468	Generic DNA sequence	4.50e+00
	23	23	20.2	431 19	T19913	Human gene signature	4.50e+00
	24	23	20.2	675 44	V11706	Carnobacteriocin BM1	4.50e+00
	25	22	19.3	501 3	N50025	Sequence encoding new	1.06e+01
	26	22	19.3	501 3	N50030	Sequence encoding new	1.06e+01
	27	22	19.3	501 3	N50027	Sequence encoding new	1.06e+01
c	28	22	19.3	501 3	N50026	Sequence encoding new	1.06e+01
c	29	22	19.3	501 3	N50030	Sequence encoding new	1.06e+01
c	30	22	19.3	501 3	N50023	Sequence encoding new	1.06e+01
c	31	22	19.3	501 3	N50032	Sequence encoding new	1.06e+01
c	32	22	19.3	501 3	N50024	Sequence encoding new	1.06e+01
	33	22	19.3	1308 44	V35184	Thuja plicatissima	1.06e+01
	34	22	19.3	2183 1	N90700	Clal fragment contg.	1.06e+01
	35	22	19.3	2976 17	T04687	Black widow spider de	1.06e+01
c	36	22	19.3	3440 3	Q13967	Bee-4 gene with regul	1.06e+01
c	37	22	19.3	3440 40	V15268	Brassica campestris B	1.06e+01
	38	22	19.3	3706 17	T04688	Human REST protein de	1.06e+01
	39	22	18.4	1090 24	T41410	Human neuron restrict	2.46e+01
	40	21	18.4	2043 23	T36898	Human REST protein co	2.46e+01
	41	21	18.4	3291 24	T41405	Insert of PSRQ220 con	2.46e+01
	42	21	18.4	5595 4	Q25810	NC-contig derived fro	2.46e+01
	43	21	18.4	80240 59	V83940	Mycoplasma genitalium	2.46e+01
	44	21	18.4	580073 27	T58840		

ALIGNMENTS

RESULT 1
ID T67141 standard; DNA; 114 BP.
AC T67141;
DE 10-DEC-1997 (first entry)
KW IF gene for bacteriocin-inducing peptide.
KW bacteriocin; vaccine; promoter; lactic acid bacterium;
OS Lactobacillus sake strain LTH 673.
FH Key Location/Qualifiers
FT signal_peptide 1..54 /*tag= a
FT mat_peptide 55..111 /*tag= b
FT /*product= mature IF peptide
FN WO9718316-A1.
PD 22-MAY-1997.
PF 13-NOV-1996; NO0266.
PR 13-NOV-1995; NO-004575.
PA (BRUR/) BRURBERG M B.
PA (EJJS/) EJJSINK V G H.
PA (NESI/) NES I F.
PI Brurberg MB, Ejjsink VGH, Nes IF;
DR P-PSDB; W17988.
PT Gene expression system providing regulated lactic acid bacteria
PT protein production uses new inducing peptide involved in
PT bacteriocin synthesis, useful in fermentation and as a drug delivery
PT system
PS Example 4; Fig 2; 39pp; English.
CC This DNA sequence comprises the IF gene of Lactobacillus sake
CC LTH673 that encodes a secreted peptide (see W17988) which
CC autoinduces the expression of the IF-K-R gene cluster involved in
CC sakacin C production. The invention relates to the discovery of a
CC new regulatory mechanism for gene expression in lactic acid

CC bacteria that includes previously unrecognized, strongly
CC regulatable promoter elements (see T6142-48). Expression of genes
CC under the control of such promoters is dependent upon expression of
CC the IF-K-R gene cluster or an analogue. Thus, the expression of a
CC gene under the control of the promoter element in cells that
CC contain the K and R genes can be induced at will by adding the
CC cognate, synthetic (harmless) inducing peptide. A recombinant
CC vector containing the gene expression system, and host cells
CC transformed with this vector, incorporated in the chromosome,
CC and/or having integrated into its chromosome a promoter linked to
CC an integrated gene encoding a specific protein are claimed.
CC Typical applications are in fermentations (e.g. where the gene for
CC an enzyme is regulated) and production of specific proteins, or
CC where the bacteria express a surface antigen, as vaccines.
SQ Sequence 114 BP; 50 A; 12 C; 16 G; 36 T;

Query Match 100.0%; Score 114; DB 34; Length 114;
Best Local Similarity 100.0%; Pred. No. 4.71e-41;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 atgatgatatttataaaactttcagaaaagaattgcataaaataaacggtggtatgcca 60
QY 1 ATGATGATATTAAAAAATTTTCAGAAAAAGAAATTCAGAAAAATAAACGGTGTATGGCA 60
|||||
Db 61 ggaataagtctattttatcataagattaaacaaattttaccatcgtaa 114
QY 61 GGAATAAGTCTCTTAATTTATTCATAGATTAAACAAATTTTACCATCGTTAA 114
|||||

RESULT 2
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI; 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 26.3%; Score 30; DB 9; Length 91;
Best Local Similarity 0.0%; Pred. No. 8.34e-03;
Matches 0; Conservative 39; Mismatches 9; Indels 0; Gaps 0;

Db 12 svhsyvvhvshhshvshvhhvshvsvvvhvshvhhvshvhhvshv 59
QY 3 GATGATATTAAAAAATTTTCAGAAAAAGAAATTCAGAAAAATAAACGG 50
|||||

RESULT 3
ID V44650 standard; DNA; 91 BP.
AC V44650;
DT 06-OCT-1998 (first entry)
DE Mammalian DNA replication origin consensus sequence, uniorconsensus.
KW DNA replication origin; human; mammal; alphaconsensus; uniorconsensus;
KW anti-gene; DNA replication inhibitor; shuttle vector construct creation;
KW gene therapy; ss.
OS Mammalia.

PN WO9827200-A2.
PD 25-JUN-1998.
PF 12-DEC-1997; CA0972.
PR 21-MAY-1997; US-047322.
PR 16-DEC-1996; US-033374.
PA (UYMC-) UNIV MCGILL.
PI Cossens NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;
PI WPI; 98-362770/31.
DR Human or mammalian origin of replication consensus sequences - for
PT inhibiting DNA replication, for controlling initiation of
PT replication, maintaining circular plasmids and in assembly of human
PT artificial chromosomes
PS Claim 1; Page 42; 54pp; English.
CC This sequence represents a human or mammalian DNA replication origin
CC consensus sequences of the invention, designated uniorconsensus.
CC Administration of the consensus sequence or an anti-gene (comprising a
CC double stranded copy of the consensus) is used to inhibit DNA replication
CC in vivo or in vitro. The consensus sequences can also be inserted into an
CC expression vector, used subsequently for in vitro transfection of
CC mammalian cells, to control initiation of DNA replication. They can also
CC be used to maintain circular plasmids that are capable of
CC semi-conservative replication in proliferating mammalian cells, or
CC inserted into mammalian or human artificial chromosome vectors for gene
CC therapy. Particularly, they are used to create shuttle vector constructs
CC for defining the essential mammalian elements required for maintenance of
CC chromosomal function. The consensus sequence can be combined with cloned
CC human telomeres and large centromeric blocks for assembly of human
CC artificial chromosomes and maintained as bacterial plasmids, circular or
CC linear, large or small yeast artificial chromosomes (YACs) or as episomal
CC elements.
SQ Sequence 91 BP; 15 A; 1 C; 4 G; 7 T;

Query Match 26.3%; Score 30; DB 46; Length 91;
Best Local Similarity 13.6%; Pred. No. 8.34e-03;
Matches 12; Conservative 47; Mismatches 29; Indels 0; Gaps 0;

Db 4 twakrawwwkkdvvwgakrwwkvvhrassacmckwaaktwkggtwarrywkrkm 63
Cp 103 TAAAAATTTGTTTATCTTATCTTATGAATAAAATAGAACTATTCTCTGCCATACCACCGTTTA 44
|||||
Db 64 wwtkawsdatakwkwwkdakwkmwrktt 91
Cp 43 TTTTGTGCAATCTTTTCTGAAAGTTT 16
|||||

RESULT 4
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI; 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 26.3%; Score 30; DB 9; Length 91;
Best Local Similarity 0.0%; Pred. No. 8.34e-03;

30-MAR-1988; 105163.
03-APR-1987; US-034819.
(SUSO) SUOMEN SOKERI OY.
Lehtovaara P, Kowles J, Koivula A, Bamford J, Reinikainen T;
WPI: 88-279927/40.
Introducing random point mutations into nucleic acids -
by prep'n of single stranded template, annealing a primer, elongation,
misincorporation, completion of molecules and screening.
Disclosure; p: English.
Random point mutations were introduced into the alpha fragment of
E.coli beta-galactosidase. The wild type sequence was obtained as a
single stranded template and an oligonucleotide was hybridised to
it to generate a pop'n of DNA molecules which terminate at all
possible nucleotide positions within a specified region. The
variable 3' ends generated in this way are used as primers for
reverse transcriptase. Nucleotides are misincorporated by the
transcriptase and the molecules are completed to forms that can be
amplified and then expressed in a suitable host-vector system.
The sequence covers all 176 diff base substitutions, most of which
occurred singularly in any given mutant.
See also P80575.
Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 25.4%; Score 29; DB 1: Length 204;
Best Local Similarity 10.1%; Pred.No. 2.10e-02;
Matches 10; Conservative 47; Mismatches 42; Indels 0; Gaps

87 mrttthymrbnvrdynrsdaaawccyrsvkydcynachdhvvybbvynvh 146
100 AAATTTCTTTAAATCTTATGAATAAATTAGACTATTCTCGCATACCACGGTTATT 41
147 nhnncncbnnhcvnhvbnhrnwayrvrhdarddhv 185
40 TTGTCAATCTTTCTTGAAAGTTTTTAAATATCATCA 2

RESULT 7
ID N50024 standard; DNA; 501 BP.
AC N50024;
04-SEP-1991 (first entry)
Sequence encoding new modified human beta interferon polypeptides
DE IFNX 417.
KW Antiviral; cell growth regulator; immune system regulator;
KW antiproliferative; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..501
FT /*tag= a
FP-163993-A.
PD 11-DEC-1985.
PF 17-MAY-1985; 105750.
PR 17-MAY-1984; GB-012564.
PA (SEAR) SEARLE G & CO.
PI Bell LD, Boseley PG, Porter AG;
WPI: 85-311944/50.
P-PSDB; P50023.
DR New modified human beta interferon polypeptide(s) - prep'd. by
PT plasmid transformed bacteria, with improved antiviral,
PT anti-proliferative and immune regulating actions
PT Claim 28; Chart 2b, page 33; 71pp; English.
PS Compared with interferon beta prep'd. by recombinant methods, the
CC INFs of the invention are more active and have different affinities
CC for cell surface receptors (allowing selective targeting); they
CC have higher therapeutic index; improved stability against microbial
CC breakdown during synthesis; and better in vivo solubility and
CC stability. they are also easier to recover from incubation mixts.
CC Sequence 501 BP; 110 A; 32 C; 66 G; 81 T;

Query Match 24.6%; Score 28; DB 3: Length 501;
Best Local Similarity 44.6%; Pred.No. 5.26e-02;
Matches 25; Conservative 15; Mismatches 16; Indels 0; Gaps

366 baarmgdtaytaygnmgdathytbcyatyvttbaatgcnaargatcayvncatt 421

44 TAAACGGTGGTATGCCAGGAAATAGTCTAAATTTTATCATAGATTAAACAAATT 99

8
RESULT V44650 standard; DNA; 91 BP.
V44650;
06-OCT-1998 (first entry)
Mammalian DNA replication origin consensus sequence, uniorconsensus.
DNA replication origin; human; mammal; alphaconsensus; uniorconsensus;
anti-gene; DNA replication inhibitor; shuttle vector construct creation;
gene therapy; ss.
Mammalia.
COS W09827200-A2.
PD 25-JUN-1998.
PPF 12-DEC-1997; CA0972.
PPR 21-MAY-1997; US-047322.
PPR 16-DEC-1996; US-033374.
PPA (UYMC-) UNIV MCGILL.
COSCONS NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;
WPI; 98-362770/31.
DR Human or mammalian origin of replication consensus sequences - for
inhibiting DNA replication, for controlling initiation of
replication maintaining circular plasmids and in assembly of human
artificial chromosomes
PT Claim 1; Page 42; 54pp; English.
PS This sequence represents a human or mammalian DNA replication origin
CC consensus sequences of the invention, designated uniorconsensus.
CC Administration of the consensus sequence or an anti-gene (comprising a
CC double stranded copy of the consensus) is used to inhibit DNA replication
CC in vivo or in vitro. The consensus sequences can also be inserted into a
CC expression vector, used subsequently for in vitro transfection of
CC mammalian cells, to control initiation of DNA replication. They can also
CC be used to maintain circular plasmids that are capable of
CC semi-conservative replication in proliferating mammalian cells, or
CC inserted into mammalian or human artificial chromosome vectors for gene
CC therapy. Particularly, they are used to create shuttle vector constructs
CC for defining the essential mammalian elements required for maintenance of
CC chromosomal function. The consensus sequence can be combined with cloned
CC human telomeres and large centromeric blocks for assembly of human
CC artificial chromosomes and maintained as bacterial plasmids, circular
CC linear, large or small yeast artificial chromosomes (YACs) or as episomal
CC elements.
SO sequence 91 BP: 15 A: 1 C: 4 G: 7 T:

17-MAY-1984; GB-012564.
(SEAR) (SEARLE G D & CO.
Bell LD, Boseley PG, Porter AG;
WPI: 85-311944/50.
P-PSDB: P50022.
New modified human beta interferon polypeptide(s) - prepd. by
plasmid transformed bacteria, with improved antiviral,
anti-proliferative and immune regulating actions
Claim 28; Chart 2a, page 32; 71pp: English.
Compared with interferon beta prepd. by recombinant methods, the
INFS of the invention are more active and have different affinities
for cell surface receptors (allowing selective targeting); they
have higher therapeutic index; improved stability against microbial
breakdown during synthesis; and better in vivo solubility and
stability. They are also easier to recover from incubation mixts.
Sequence 501 BP; 107 A; 31 C; 69 G; 80 T;

RESULT	10	
ID	N50026 standard; DNA; 501 BP.	
AC	N50026;	
DE	40-SEP-1991 (first entry)	
DT	Sequence encoding new modified human beta interferon polypeptides	
DE	IFNX 430.	
DE	Antiviral; cell growth regulator; immune system regulator;	
KW	antiproliferative; ss.	
KW	Homo sapiens.	
OS		
FH	Key	Location/Qualifiers
ET	1..501	
FT	/*tag= a	
FP	EP-163993-A.	
PD	11-DEC-1985.	
PF	17-MAY-1985; 105750.	
PR	17-MAY-1984; GB-012564.	
PR	(SEAR) SEARLE G D & CO.	
PI	Bell LD, Boseley PG, Porter AG;	
PI	WPI: 85-311944/50.	
DR	P-PSDB; P50025.	
PT	New modified human beta interferon polypeptide(s) - prepd. by	
PT	plasmid transformed bacteria, with improved antiviral,	
PT	anti-proliferative and immune regulating actions	
PS	Claim 28; Chart 2d, page 35; 71pp; English.	
CC	Compared with interferon beta prepd. by recombinant methods, the	
CC	INFS of the invention are more active and have different affinities	
CC	for cell surface receptors (allowing selective targeting); they	
CC	have higher therapeutic index; improved stability against microbial	
CC	breakdown during synthesis; and better in vivo solubility and	
CC	stability. They are also easier to recover from incubation mixts.	
SC	sequence, 501 Bp; 108 A; 31 C; 70 G; 81 T;	

RESULT 11
ID Q70457 standard; DNA; 114 BP.
AC Q70467;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
DE TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;

RESULT	9	
ID	N50023	standard; DNA; 501 BP.
AC	N50023:	
DT	04-SEP-1991	(first entry)
DE	Sequence encoding new modified human beta interferon polypeptides	
DE	IFNX 416.	
DE	Antiviral; cell growth regulator; immune system regulator;	
KW	antiproliferative; ss.	
KW	Homo sapiens.	
OS	Location/Qualifiers	
FT	Key	
FT	1..501	
FT	/*tag= a	
PN	EP-163993-A.	
PD	11-DEC-1985.	
PF	17-MAY-1985.	105750.

KW effector domain; concatenated heterofunctional protein; linker;
 OS direct; rapid; detection; screening; treatment; generic; ss.
 FH Key
 FT misc_feature
 FT 55..60
 FT Location/Qualifiers
 FT /*tag= a
 FT /note= "this sequence represents 'z'; 2 can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 FT
 FT W09418318-A.
 PD 18-AUG-1994.
 PE 01-FEB-1994; U00977.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UYNC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 DR P-PSDB; R65153.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC Q70467 is a generic DNA sequence used to generate random TSAR (Totally
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be
 CC represented as follows: X(NNB)16(TGC)(NNB)12(NNB)16(TGC)(NNB)12. X
 CC and Y are flanking restriction sites (X is not the same as Y) that are
 CC not specified further. Other generic sequences are shown in Q70466-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC Q65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compsns. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need for
 CC complex methods of hybridoma formation or in vivo antibody production.
 CC The TSARs are easily characterised and have designed activity allowing
 CC direct and rapid detection in a screening process.
 SQ Sequence 114 BP; 2 C; 2 G; 2 T;
 0 A; 2 C; 2 G; 2 T;
 Query Match 21.1%; Score 24; DB 12; Length 114;
 Best Local Similarity 2.8%; Pred. No. 1.89e+00;
 Matches 3; Conservative 27; Mismatches 76; Indels 0; Gaps 0;
 Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 62
 Cp 106 GGGTAAATTTGTTTAACTTATGTAATAAATAGACTATTTCTGCAATACCACCGT 47
 Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 108
 Cp 46 TTATTTTTCGAATCTTTTCTGAAAGTTTAAATATCATCAT 1
 RESULT 12
 ID N50034 standard; DNA; 498 BP.
 AC N50034;
 DT 04-SEP-1991 (first entry)
 DE Sequence encoding new modified human beta interferon polypeptides
 DE IFN4 485.
 KW Antiviral; cell growth regulator; immune system regulator;
 KW antiproliferative; ss.
 OS Homo sapiens.
 FH Key
 FT Cds
 FT 1..498
 FT Location/Qualifiers
 FT /*tag= a
 PN EP-163993-A.

PD 11-DEC-1985.
 PF 17-MAY-1985; 105750.
 PR 17-MAY-1984; GB-012564.
 PA (SEAR) SEARLE G D & CO.
 PI Bell LD, Boseley PG, Porter AG;
 DR WPI: 85-311944/50.
 DR P-PSDB; P50033.
 PT New modified human beta interferon polypeptide(s) - prepd. by
 PT plasmid transformed bacteria, with improved antiviral,
 PT anti-proliferative and immune regulating actions
 PS Claim 28; Chart 21, page 43; 71pp; English.
 CC Compared with interferon beta prep. by recombinant methods, the
 CC INFs of the invention are more active and have different affinities
 CC for cell surface receptors (allowing selective targeting); they
 CC have higher therapeutic index; improved stability against microbial
 CC breakdown during synthesis; and better in vivo solubility and
 CC stability. They are also easier to recover from incubation mixts.
 SQ Sequence 498 BP; 112 A; 30 C; 68 G; 77 T;
 Query Match 21.1%; Score 24; DB 3; Length 498;
 Best Local Similarity 39.3%; Pred. No. 1.89e+00;
 Matches 22; Conservative 16; Mismatches 18; Indels 0; Gaps 0;
 Db 363 baarmgdtaytvggmngdathytbcaytaytbaargnaargartaywnncayt 418
 Qy 44 TAAACGGTGTATGGCAGGAATAGTCTAATTTTATTCATAAGATTAAACAAATT 99
 RESULT 13
 ID N50032 standard; DNA; 501 BP.
 AC N50032;
 DT 04-SEP-1991 (first entry)
 DE Sequence encoding new modified human beta interferon polypeptides
 DE IFN4 449.
 KW Antiviral; cell growth regulator; immune system regulator;
 KW antiproliferative; ss.
 OS Homo sapiens.
 FH Key
 FT Cds
 FT 1..501
 FT Location/Qualifiers
 FT /*tag= a
 PN EP-163993-A.
 PD 11-DEC-1985.
 PR 17-MAY-1985; 105750.
 PR 17-MAY-1984; GB-012564.
 PA (SEAR) SEARLE G D & CO.
 PI Bell LD, Boseley PG, Porter AG;
 DR WPI: 85-311944/50.
 DR P-PSDB; P50031.
 PT New modified human beta interferon polypeptide(s) - prepd. by
 PT plasmid transformed bacteria, with improved antiviral,
 PT anti-proliferative and immune regulating actions
 PS Claim 28; Chart 21, page 41; 71pp; English.
 CC Compared with interferon beta prep. by recombinant methods, the
 CC INFs of the invention are more active and have different affinities
 CC for cell surface receptors (allowing selective targeting); they
 CC have higher therapeutic index; improved stability against microbial
 CC breakdown during synthesis; and better in vivo solubility and
 CC stability. They are also easier to recover from incubation mixts.
 SQ Sequence 501 BP; 108 A; 30 C; 72 G; 79 T;
 Query Match 21.1%; Score 24; DB 3; Length 501;
 Best Local Similarity 39.3%; Pred. No. 1.89e+00;
 Matches 22; Conservative 16; Mismatches 18; Indels 0; Gaps 0;
 Db 366 baarmgdtaytvggmngdathytbcaytaytbaargnaargartaywnncayt 421
 Qy 44 TAAACGGTGTATGGCAGGAATAGTCTAATTTTATTCATAAGATTAAACAAATT 99
 RESULT 14
 ID N50031 standard; DNA; 501 BP.
 AC N50031;
 DT 04-SEP-1991 (first entry)

W P S R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Wed Sep 15 10:38:38 1999; MasPar time 9.71 Seconds
Tabular output not generated. 831.838 Million cell updates/sec

Title: >US-09-068-507A-2
Description: (1-114) from US09068507A.seq
Perfect Score: 114
N.A. Sequence: 1 ATGATGATATTAAAAACT.....AAATTTTACCACGTTAA 114
Comp: TACTACTATAAATTTTGA.....TTTAAATGGTAGCAATT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 137068 seqs, 95432894 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-issued

1:5A_COMB 2:5B_COMB 3:5C_COMB 4:PCT9_COMB 5:backfiles1

Statistics: Mean 6.588; Variance 4.618; scale 1.427

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	27	23.7	7218	2	US-08-232-Sequence 14, Applicatio	1.41e-03
2	22	19.3	2183	1	US-07-993-Sequence 1, Applicatio	3.36e-01
3	22	19.3	3440	4	PCT-US91-0-Sequence 27, Applicati	3.36e-01
4	22	19.3	3440	2	US-08-471-Sequence 27, Applicati	3.36e-01
5	21	18.4	928	4	PCT-US96-0-Sequence 5, Applicatio	9.62e-01
6	21	18.4	1090	4	PCT-US96-0-Sequence 4, Applicatio	9.62e-01
7	21	18.4	1407	4	PCT-US96-0-Sequence 3, Applicatio	9.62e-01
8	21	18.4	1461	4	PCT-US96-0-Sequence 27, Applicati	9.62e-01
9	21	18.4	1791	4	PCT-US96-0-Sequence 10, Applicatio	9.62e-01
10	21	18.4	3291	4	PCT-US96-0-Sequence 1, Applicatio	9.62e-01
11	21	18.4	5595	1	US-07-635-Sequence 1, Applicatio	9.62e-01
12	21	18.4	5595	1	US-07-841-Sequence 1, Applicatio	9.62e-01
13	21	18.4	5648	1	US-08-238-Sequence 1, Applicatio	9.62e-01
14	20	17.5	215	1	US-08-238-Sequence 5, Applicatio	2.71e+00
15	20	17.5	729	1	US-08-189-Sequence 9, Applicatio	2.71e+00
16	20	17.5	729	3	US-08-188-Sequence 26, Applicati	2.71e+00
17	20	17.5	965	3	US-08-388-Sequence 22, Applicati	2.71e+00
18	20	17.5	1670	3	US-08-188-Sequence 5, Applicatio	2.71e+00
19	20	17.5	1670	1	US-08-189-Sequence 10, Applicati	2.71e+00
20	20	17.5	1670	1	US-08-189-Sequence 10, Applicati	2.71e+00

21	20	17.5	1969	3	US-08-188-Sequence 3, Applicatio	2.71e+00
22	20	17.5	2687	1	US-08-149-Sequence 8, Applicatio	2.71e+00
23	20	17.5	2687	1	US-08-377-Sequence 8, Applicatio	2.71e+00
24	20	17.5	2873	1	US-08-149-Sequence 1, Applicatio	2.71e+00
25	20	17.5	2873	1	US-08-377-Sequence 1, Applicatio	2.71e+00
26	20	17.5	2913	3	US-08-188-Sequence 6, Applicatio	2.71e+00
27	20	17.5	3180	4	PCT-US96-0-Sequence 1, Applicatio	2.71e+00
28	20	17.5	3180	3	US-08-480-Sequence 1, Applicatio	2.71e+00
29	19	16.7	1145	5	5510472-1 Patent No. 5510472	2.71e+00
30	19	16.7	1288	3	US-08-440-Sequence 9, Applicatio	2.71e+00
31	19	16.7	2051	3	US-08-482-Sequence 2, Applicatio	2.71e+00
32	19	16.7	2975	2	US-08-368-Sequence 1, Applicatio	2.71e+00
33	19	16.7	3033	3	US-08-482-Sequence 1, Applicatio	2.71e+00
34	19	16.7	3240	2	US-08-368-Sequence 3, Applicatio	2.71e+00
35	19	16.7	3303	1	US-08-081-Sequence 3, Applicatio	2.71e+00
36	18	15.8	198	4	PCT-US92-1-Sequence 16, Applicati	2.71e+00
37	18	15.8	660	1	US-07-991-Sequence 32, Applicati	2.02e+01
38	18	15.8	1511	1	US-07-991-Sequence 8, Applicatio	2.02e+01
39	18	15.8	1511	2	US-08-107-Sequence 8, Applicatio	2.02e+01
40	18	15.8	2306	4	PCT-US96-0-Sequence 3, Applicatio	2.02e+01
41	18	15.8	2760	1	US-08-101-Sequence 1, Applicatio	2.02e+01
42	18	15.8	2760	1	US-08-101-Sequence 3, Applicatio	2.02e+01
43	18	15.8	4837	1	US-08-089-Sequence 1, Applicatio	2.02e+01
44	18	15.8	8457	1	US-07-991-Sequence 1, Applicatio	2.02e+01
45	18	15.8	9636	3	US-08-323-Sequence 1, Applicatio	2.02e+01

ALIGNMENTS

RESULT 1
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC xxxxxx
DT
DE Sequence 14, Application US/08232463
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
CC APPLICANT: DORNER, F.
CC APPLICANT: SCHEIFLINGER, F.
CC APPLICANT: FALKNER, F. G.
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
CC NUMBER OF SEQUENCES: 52
CC CORRESPONDENCE ADDRESS:
CC ADDRESSES: Foley & Lardner
CC STREET: 1800 Diagonal Road, Suite 500
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22313-0299
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/232.463
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/07/935.313
CC FILING DATE:
CC APPLICATION NUMBER: EP 91 114 300.6
CC FILING DATE: 26-AUG-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BENT, Stephen A.
CC REGISTRATION NUMBER: 29,768
CC REFERENCE/DOCKET NUMBER: 30472/114 IMM
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)836-9300
CC TELEFAX: (703)683-4109
CC TELEX: 899149
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:

一

CC	TITLE OF INVENTION: REST Protein and DNA
CC	NUMBER OF SEQUENCES: 29
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: Dechert Price & Rhoads
CC	STREET: P.O. Box 5218
CC	CITY: Princeton
CC	STATE: New Jersey
CC	COUNTRY: USA
CC	ZIP: 08543-5218
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC	COMPUTER: IBM-compatible
CC	OPERATING SYSTEM: DOS 5.0
CC	SOFTWARE: WordPerfect
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: PCT/US96/03940
CC	FILING DATE: March 23, 1995
CC	CLASSIFICATION:
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: Allen Bloom
CC	REGISTRATION NUMBER: 29,135
CC	REFERENCE/DOCKET NUMBER: 317743-101
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: (609) 520-3214
CC	TELEFAX: (609) 520-3259
CC	INFORMATION FOR SEQ ID NO: 5:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 928 base pairs
CC	TYPE: nucleic acid
CC	STRANDEDNESS: double
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: cDNA to mRNA
CC	HYPOTHETICAL: no
CC	ANTI-SENSE: no
CC	ORIGINAL SOURCE:
CC	ORGANISM: Human
CC	CELL LINE: HeLa
CC	IMMEDIATE SOURCE:
CC	LIBRARY: cDNA
CC	PUBLICATION INFORMATION:
CC	AUTHORS: Chong, Jayhong A., Tapia-Ramirez Jos , Toledo-
CC	AUTHORS: Aral, Juan, Zheng, Yingcong, Boutros, Michael C
CC	AUTHORS: Yelena M., Frohman, Michael A., Kraner, Susan D
CC	TITLE: REST: A Mammalian Silencer Protein that Restricts
CC	JOURNAL: Cell
CC	VOLUME: 80
CC	ISSUE:
CC	PAGES:
CC	DATE: March 24, 1995
CC	RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 928
CC	SEQUENCE 928 BP; 363 A; 178 C; 182 G; 205 T; 0 OTHER.
CC	Query Match 18.4%; Score 21; DB 4; Length 928;
CC	Best Local Similarity 74.4%; Pred. No. 9.62e-01;
CC	Matches 32; Conservative 0; Mismatches 11; Indels 0;
DB	543 CAGAAATAGACAAACAAAATAAAAGGGGATGGCTGGGAAA 685
QY	23 CAGAAAAGAATTGCAAAAAATAAACGGTGTATGGCAGAAA 65
CC	
CC	RESULT 6
ID	PCT-US96-03940-4 STANDARD; DNA; UNC; 1090 BP.
CC	xxxxxx
DT	
DE	Sequence 4, Application PC/TUS9603940
CC	Sequence 4, Application PC/TUS9603940
CC	GENERAL INFORMATION:
CC	APPLICANT: Mandel, Gall, Chong, Jayhong A.
CC	TITLE OF INVENTION: REST Protein and DNA
CC	NUMBER OF SEQUENCES: 29
CC	CORRESPONDENCE ADDRESS:

```
CC ADDRESSEE: Dechert Price & Rhoads
CC STREET: P.O. Box 5218
CC CITY: Princeton
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 08543-5218
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC COMPUTER: IBM-compatible
CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: WordPerfect
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03940
CC FILING DATE: March 23, 1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Allen Bloom
CC REGISTRATION NUMBER: 29,135
CC REFERENCE/DOCKET NUMBER: 317743-101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (609) 520-3214
CC TELEFAX: (609) 520-3259
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1090 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA to mRNA
CC HYPOTHETICAL: no
CC ANTI-SENSE: no
CC ORGANISM: Human
CC CELL LINE: HeLa
CC IMMEDIATE SOURCE:
CC LIBRARY: cDNA
CC PUBLICATION INFORMATION:
CC AUTHORS: Chong, Jayhong A., Tapia-Ramirez Jos , Toledo-
CC AUTHORS: Aral, Juan, Zheng, Yingcong, Boutros, Michael C., Altschuler,
CC AUTHORS: Yelena M., Frohman, Michael A., Kraner, Susan D., Mandel, Gail
CC TITLE: REST: A Mammalian Silencer Protein that Restricts
CC JOURNAL: Cell
CC VOLUME: 80
CC DATE: March 24, 1995
CC PAGES:
CC RELEVANT RESIDUES IN SEQ ID NO: 4: FROM 1 TO 1090
CC SEQUENCE 1090 BP; 433 A; 206 C; 219 G; 232 T; 0 OTHER.
CC
CC Query Match 18.4%; Score 21; DB 4; Length 1090;
CC Best Local Similarity 74.4%; Pred. No. 9.62e-01;
CC Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
CC
CC Db 702 CAGAAATAGAACACAAATAAAAGGGGATGGCTGGAAA 744
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC Qy 23 CAGAAAAGAAATTGCAAAAATAAAGCGTGTATGCGCAGGAAA 65
CC
CC RESULT 7
CC ID PCT-US96-03940-3 STANDARD; DNA; UNC; 1407 BP.
CC AC xxxxxx
CC DT
CC DE Sequence 3, Application PC/TUS9603940
CC CC Sequence 3, Application PC/TUS9603940
CC CC GENERAL INFORMATION:
CC CC APPLICANT: Mandel, Gail, Chong, Jayhong A.
CC CC TITLE OF INVENTION: REST Protein and DNA
CC CC NUMBER OF SEQUENCES: 29
CC CC CORRESPONDENCE ADDRESS:
CC CC ADDRESSEE: Dechert Price & Rhoads
CC CC STREET: P.O. Box 5218
CC CC CITY: Princeton
CC
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 08543-5218
CC
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 08543-5218
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC COMPUTER: IBM-compatible
CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: WordPerfect
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03940
CC FILING DATE: March 23, 1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Allen Bloom
CC REGISTRATION NUMBER: 29,135
CC REFERENCE/DOCKET NUMBER: 317743-101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (609) 520-3214
CC TELEFAX: (609) 520-3259
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1407 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA to mRNA
CC HYPOTHETICAL: no
CC ANTI-SENSE: no
CC ORGANISM: Human
CC CELL LINE: HeLa
CC IMMEDIATE SOURCE:
CC LIBRARY: cDNA
CC PUBLICATION INFORMATION:
CC AUTHORS: Chong, Jayhong A., Tapia-Ramirez Jos , Toledo-
CC AUTHORS: Aral, Juan, Zheng, Yingcong, Boutros, Michael C., Altschuler,
CC AUTHORS: Yelena M., Frohman, Michael A., Kraner, Susan D., Mandel, Gail
CC TITLE: REST: A Mammalian Silencer Protein that Restricts
CC JOURNAL: Cell
CC VOLUME: 80
CC DATE: March 24, 1995
CC PAGES:
CC RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 1 TO 1407
CC SEQUENCE 1407 BP; 515 A; 276 C; 306 G; 310 T; 0 OTHER.
CC
CC Query Match 18.4%; Score 21; DB 4; Length 1407;
CC Best Local Similarity 74.4%; Pred. No. 9.62e-01;
CC Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
CC
CC Db 1122 CAGAAATAGAACACAAATAAAAGGGGATGGCTGGAAA 1164
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC Qy 23 CAGAAAAGAAATTGCAAAAATAAAGCGTGTATGCGCAGGAAA 65
CC
CC RESULT 8
CC ID PCT-US96-03940-27 STANDARD; DNA; UNC; 1461 BP.
CC AC xxxxxx
CC DT
CC DE Sequence 27, Application PC/TUS9603940
CC CC Sequence 27, Application PC/TUS9603940
CC CC GENERAL INFORMATION:
CC CC APPLICANT: Mandel, Gail, Chong, Jayhong A.
CC CC TITLE OF INVENTION: REST Protein and DNA
CC CC NUMBER OF SEQUENCES: 29
CC CC CORRESPONDENCE ADDRESS:
CC CC ADDRESSEE: Dechert Price & Rhoads
CC CC STREET: P.O. Box 5218
CC CC CITY: Princeton
CC
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 08543-5218
```



```

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC COMPUTER: IBM-compatible
CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: WordPerfect
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03940
CC FILING DATE: March 23, 1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Allen Bloom
CC REGISTRATION NUMBER: 29,135
CC REFERENCE/DOCKET NUMBER: 317743-101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (609) 520-3214
CC TELEFAX: (609) 520-3259
CC INFORMATION FOR SEQ ID NO: 27:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1461 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cdna to mrna
CC HYPOTHETICAL: no
CC ANTI-SENSE: no
CC ORIGINAL SOURCE:
CC ORGANISM: Human
CC CELL LINE: HeLa
CC IMMEDIATE SOURCE:
CC LIBRARY: CDNA
CC PUBLICATION INFORMATION:
CC AUTHORS: Chong, Jayhong A., Tapia-Ramirez Jos , Toledo-
CC AUTHORS: Aral, Juan, Zheng, Yingcong, Boutros, Michael C., Altschuler,
CC AUTHORS: Yelena M., Frohman, Michael A., Kraner, Susan D., Mandel, Gail
CC TITLE: REST: A Mammalian Silencer Protein that Restricts
CC TITLE: Sodium Channel Gene Expression to Neurons
CC JOURNAL: Cell
CC VOLUME: 80
CC ISSUE:
CC PAGES:
CC DATE: March 24, 1995
CC RELEVANT RESIDUES IN SEQ ID NO: 26:FROM 1 TO 1461
CC SEQUENCE 1461 BP: 528 A; 288 C; 326 G; 319 T; 0 OTHER.
SQ
Query Match 18.4%; Score 21; DB 4; Length 1461;
Best Local Similarity 74.4%; Pred. No. 9.62e-01;
Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Db 1217 CAGAAATAGACAAACAAATAAAGGGGATGTCGCTGGAAA 1259
||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 23 CAGAAAAGAAATTCGAAAAATAAACGGTGTGTATGGCAGGAAA 65
RESULT 9
ID PCT-US96-03940-6 STANDARD; DNA; UNC; 1791 BP.
AC xxxxxx
DE
BT
Sequence 6, Application PC/TUS9603940
CC Sequence 6, Application PC/TUS9603940
CC GENERAL INFORMATION:
CC APPLICANT: Mandel, Gail, Chong, Jayhong A.
CC TITLE OF INVENTION: REST Protein and DNA
CC NUMBER OF SEQUENCES: 29
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dechert Price & Rhoads
CC STREET: P.O. Box 5218
CC CITY: Princeton
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 08543-5218
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC COMPUTER: IBM-compatible

```

```

CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: WordPerfect
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03940
CC FILING DATE: March 23, 1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Allen Bloom
CC REGISTRATION NUMBER: 29,135
CC REFERENCE/DOCKET NUMBER: 317743-101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (609) 520-3214
CC TELEFAX: (609) 520-3259
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1791 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cdna to mrna
CC HYPOTHETICAL: no
CC ANTI-SENSE: no
CC ORIGINAL SOURCE:
CC ORGANISM: Human
CC CELL LINE: HeLa
CC IMMEDIATE SOURCE:
CC LIBRARY: CDNA
CC PUBLICATION INFORMATION:
CC AUTHORS: Chong, Jayhong A., Tapia-Ramirez Jos , Toledo-
CC AUTHORS: Aral, Juan, Zheng, Yingcong, Boutros, Michael C., Altschuler,
CC AUTHORS: Yelena M., Frohman, Michael A., Kraner, Susan D., Mandel, Gail
CC TITLE: REST: A Mammalian Silencer Protein that Restricts
CC TITLE: Sodium Channel Gene Expression to Neurons
CC JOURNAL: Cell
CC VOLUME: 80
CC ISSUE:
CC PAGES:
CC DATE: March 24, 1995
CC RELEVANT RESIDUES IN SEQ ID NO: 6:FROM 1 TO 1791
CC SEQUENCE 1791 BP: 607 A; 395 C; 406 G; 383 T; 0 OTHER.
SQ
Query Match 18.4%; Score 21; DB 4; Length 1791;
Best Local Similarity 74.4%; Pred. No. 9.62e-01;
Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Db 1512 CAGAAATAGACAAACAAATAAAGGGGATGTCGCTGGAAA 1554
||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 23 CAGAAAAGAAATTCGAAAAATAAACGGTGTGTATGGCAGGAAA 65
RESULT 10
ID PCT-US96-03940-10 STANDARD; DNA; UNC; 3291 BP.
AC xxxxxx
DE
BT
Sequence 10, Application PC/TUS9603940
CC Sequence 10, Application PC/TUS9603940
CC GENERAL INFORMATION:
CC APPLICANT: Mandel, Gail, Chong, Jayhong A.
CC TITLE OF INVENTION: REST Protein and DNA
CC NUMBER OF SEQUENCES: 29
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dechert Price & Rhoads
CC STREET: P.O. Box 5218
CC CITY: Princeton
CC STATE: New Jersey
CC COUNTRY: USA
CC ZIP: 08543-5218
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC COMPUTER: IBM-compatible
CC OPERATING SYSTEM: DOS 5.0
CC SOFTWARE: WordPerfect
CC CURRENT APPLICATION DATA:

```


ADDRESSEE: Dechert Price & Rhoads
STREET: P.O. Box 5218
CITY: Princeton
STATE: New Jersey
COUNTRY: USA
ZIP: 08543-5218
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM-compatible
OPERATING SYSTEM: DOS 5.0
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/03940
FILING DATE: March 23, 1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Allen Bloom
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: 317743-101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 520-3214
TELEFAX: (609) 520-3259
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5648 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Human
CELL LINE: HeLa
IMMEDIATE SOURCE:
LIBRARY: cDNA
PUBLICATION INFORMATION:
AUTHORS: Chong, Jayhong A., Tapia-Ramirez Jos ,
AUTHORS: Toledo-Aral,
AUTHORS: Juan, Zheng, Yingcong, Boutros, Michael C
AUTHORS: M., Frohman, Michael A., Kraner, Susan D
TITLE: REST: A Mammalian Silencer Protein that Res
TITLE: Sodium Channel Gene Expression to Neurons
JOURNAL: Cell
VOLUME: 80
ISSUE:
PAGES:
DATE: March 24, 1995
RELEVANT RESIDUES IN SEQ ID NO: 1:FROM -1 TO 5648
SEQUENCE 5648 BP; 1747 A; 1098 C; 1507 T; 2 OTHER

Query Match	18.4%	Score 21	DB 4	Length 5648:
Best Local Similarity	74.4%	Pred. No. 9,62e-01		
Matches 32	Conservative 0	Mismatches 11	Indels	
Db	1661	CAGAAATGACAAACAATAAAAGGGATGTGCTGGAAA	1703	
QY	23	CAGAAAAGCAATTGCAAAATAAAGCGTGTGATGCGAGAAA	65	

RESULT	14	
ID	US-08-238-163-5	STANDARD; DNA; UNC; 215 BP.
AC	xxxxxx	
DT		
DE	Sequence 5,	Application US/08238163
CC	Sequence 5,	Application US/08238163
CC	Patent No. 5569830	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	BENNETT, Alan
CC	APPLICANT:	LABAVITCH, John M.
CC	APPLICANT:	POWELL, Ann
CC	APPLICANT:	STOTZ, Henrik
CC	TITLE OF INVENTION: PLANT INHIBITORS OF FUNGAL	

M E S R A

(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Wed Sep 15 10:47:40 1999; MasPar time 268.21 Seconds
Tabular output not generated. 995.949 Million cell updates/sec

Title: >US-09-068-507A-2
Description: (1-114) from US09068507A.seq
Perfect Score: 114
N.A. Sequence: 1 ATGATGATATTATAAAACTAAATTTTACCCATCGTTAA 114
Comp: TACTACTATAAATTTTGA.....TTTAAATGGTAGCAATT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0
Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est11
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
34:gb_est35 35:gb_est36 36:gb_est37 37:gb_est38 38:gb_gss2
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 9.510; Variance 3.329; scale 2.857

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
C 1	31	27.2	252	17	97SN1787 Rice Immature	3.63e-08
C 2	30	26.3	2275	20	AF034173 Human mRNA (T	2.71e-07
C 3	29	25.4	252	17	97SN1787 Rice Immature	1.97e-06
C 4	25	21.9	365	24	aa47e06.xl Stanley Fro	4.01e-03
C 5	25	21.9	392	24	A1267345 Stanley Fro	4.01e-03
C 6	25	21.9	675	39	CPG0587A CpiOWAGDNL C	4.01e-03
C 7	24	21.1	356	40	AQ2683705 RPC111-69M19.TJ R	2.45e-02
C 8	24	21.1	359	18	aa47e06.xl Soares 2NDM	2.45e-02
C 9	24	21.1	423	25	AU039222 Dictyostellum	2.45e-02
C 10	24	21.1	434	40	AQ268460 RPC111-74G21.TJ R	2.45e-02

11	24	21.1	435	24	AU033574	AU033574 Dictyostellum	2.45e-02
C 12	24	21.1	530	41	AQ344881	RPC111-114117.TJ R	2.45e-02
C 13	23	20.2	133	29	AI570972	to24603.xl NCI_CGAP_U	1.44e-01
C 14	23	20.2	252	21	AA996637	MCCE3650.MLF Schlusto	1.44e-01
C 15	23	20.2	233	11	AA334468	EST38687 Embryo, 9 wee	1.44e-01
C 16	23	20.2	300	15	C59929	C59929 Yuji Kohara unp	1.44e-01
C 17	23	20.2	338	8	D32810	CELK021HYR Yuji Kohara	1.44e-01
C 18	23	20.2	349	27	D178233	tm50f12.xl NCI_CGAP_K	1.44e-01
C 19	23	20.2	352	22	AI039179	ox31e12.s1 Soares tota	1.44e-01
C 20	23	20.2	362	12	AA371866	EST83933 Parathyroid 9	1.44e-01
C 21	23	20.2	384	38	B55192	CIT-HSP-386F22.TR CIT-	1.44e-01
C 22	23	20.2	400	30	R50363	YJ59605.s1 Soares brea	1.44e-01
C 23	23	20.2	403	17	AA748479	ny01f04.s1 NCI_CGAP_G	1.44e-01
C 24	23	20.2	430	8	D20068	HUMG01040 Human promy	1.44e-01
C 25	23	20.2	441	8	D32326	CELK010G9R Yuji Kohara	1.44e-01
C 26	23	20.2	515	21	C92718	C92718 Dictyostellum d	1.44e-01
C 27	23	20.2	542	22	AI066149	TENU540 T. cruzi epim	1.44e-01
C 28	23	20.2	551	21	C92168	C92168 Dictyostellum d	1.44e-01
C 29	23	20.2	568	41	AQ372536	RPC111-113020.TJ R	1.44e-01
C 30	23	20.2	612	41	AQ375760	RPC111-150J22.TV R	1.44e-01
C 31	23	20.2	724	40	AQ269610	HS_2036.B1_G07_MR CIT	1.44e-01
C 32	23	20.2	759	40	AQ253262	HS_2046.B1_A08_MR CIT	1.44e-01
C 33	23	20.2	783	28	AI537158	PZ2_114_C01.r tumor2	1.44e-01
C 34	23	20.2	804	37	B12681	F27D1-Sp6.1 IGF Arabid	1.44e-01
C 35	23	20.2	943	37	B13335	F2K7-T7 IGF Arabidopsi	1.44e-01
C 36	22	19.3	362	31	R97401	YQ53B06.s1 Soares feta	8.04e-01
C 37	22	19.3	367	9	AA200306	mu58d10.r1 Soares mous	8.04e-01
C 38	22	19.3	396	38	AQ015142	CIT-HSP-231J14.TR CIT-	8.04e-01
C 39	22	19.3	398	31	H44951	Y069008.s1 Soares brea	8.04e-01
C 40	22	19.3	411	33	N68561	za15c07.s1 Soares feta	8.04e-01
C 41	22	19.3	468	33	N52522	YV52g10.s1 Soares feta	8.04e-01
C 42	22	19.3	468	27	AI447234	mq97h06.xl Soares mous	8.04e-01
C 43	22	19.3	528	42	AQ392828	CJTB1-E1-2544J24.TR CI	8.04e-01
C 44	22	19.3	632	44	AU030522	AU030522 Rice cDNA fro	8.04e-01
C 45	22	19.3	653	17	AA636365	GM07543.Sprime GM Dros	8.04e-01

ALIGNMENTS

1
RESULT 1 AA754459 252 bp mRNA EST 20-JAN-1998
LOCUS 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
DEFINITION CDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID 92801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
1 (bases 1 to 252)
Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
Unpublished (1998)
On Jan 14, 1998 this sequence version replaced gi:1797457.
Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.osti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bnaam@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. .252
/organism="Oryza sativa"

```

/cultivar="Milyang23"
/Note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT      5 a      21 c      12 g      35 t      179 others
ORIGIN

Query Match      27.2%; Score 31; DB 17; Length 252;
Best Local Similarity 9.5%; Pred. No. 3.63e-08;
Matches 7; Conservative 41; Mismatches 26; Indels 0; Gaps 0;

Db 58 VNWSTWGTGTVNBNVSGDWHYWBVNTKVDVGNTRCSRWBVTRMAHYHDYNCBYYN 117
Cp 99 AATTGTTTAAATCTTATGAATAAATAGAACTATTCTCGCATACACCGTTTATTTT 40
Db 118 NNDYHMHBBMYBB 131
Cp 39 TTGCAATCTTTT 26

RESULT 2 AF034173 2275 bp mRNA EST 30-MAR-1998
LOCUS AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
DEFINITION clone ntcon2 contig, mRNA sequence.
ACCESSION AF034173
NID 92707735
VERSION AF034173.1 GI:27077735
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2275)
AUTHORS Tripodis,N. and Ragoussis,J.
TITLE Generation of a transcription map in the region immediately
centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
boundary
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:20451115.

Contact: Tripodis, Nikos
Division of Medical and Molecular Genetics
Guys Hospital
7th floor, Guy's Tower, London SE1 9RT, UK
Email: nikos@ki.ni.

FEATURES
Source
1..2275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6p21.3"
/clone="ntcon2 contig"
/clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      438 a      619 c      470 g      599 t      149 others
ORIGIN

Query Match      26.3%; Score 30; DB 20; Length 2275;
Best Local Similarity 12.2%; Pred. No. 2.71e-07;
Matches 11; Conservative 49; Mismatches 30; Indels 0; Gaps 0;

Db 1477 AMRYRKRWRKRGRRKMTGMYKMYRMMAMCMMACWYWKMRGKKCKWKYKYYK 1536
Cp 90 AATCTTATGAATAAATAGAACTATTCTCGCATACACCGTTTATTTTGTGAATTC 31
Db 1537 YTSYKYSRWRWYTYTYTYWCWCCTSMK 1566
Cp 30 TTTTCTGAAGTTTTTTAAATATCATCAT 1

```

```

3
RESULT 3 AA754459 252 bp mRNA EST 20-JAN-1998
LOCUS 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
DEFINITION cDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID 92801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.

```

```

Contact: Eun M.Y.
Department of Cyogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bnhnm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.

```

```

FEATURES
Source
1..252
/organism="Oryza sativa"
/cultivar="Milyang23"
/Note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT      5 a      21 c      12 g      35 t      179 others
ORIGIN

Query Match      25.4%; Score 29; DB 17; Length 252;
Best Local Similarity 12.7%; Pred. No. 1.97e-06;
Matches 14; Conservative 49; Mismatches 47; Indels 0; Gaps 0;

Db 16 BAWMTTSYBCHGNBVVVCVASHGNYMSVHNCBTRGTHCDCKNNVNSTWGTWGVNBNVSG 75
Cp 3 GATCATATTTAAAAAACTTTCAGAAAAAGAAATTCAGAAAAATAACGGTGTATGCAGG 62
Db 76 DWHYWBVNTKVDVGNTRCSRWBVTRMAHYHDYNCBYYNNDYHMH 125
Cp 63 AAATAGTCTTAATTTTATTATTAAGATTAAACAAATTTTACCATCGTT 112

```

```

4
RESULT 4 AI267433 365 bp mRNA EST 17-NOV-1998
LOCUS sq47e06.x1 Stanley Frontal NB pool 2 Homo sapiens cDNA clone
DEFINITION IMAGE:2034082, mRNA sequence.
ACCESSION AI267433
NID 93886600
VERSION AI267433.1 GI:38866600
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```


LOCUS	AQ268460	434 bp	DNA	GSS	26-OCT-1998							
DEFINITION	RPC111-74G21.TJ RPC111 Homo sapiens genomic clone R-74G21, genomic survey sequence.											
ACCESSION	AQ268460											
NID	53796064											
VERSION	AQ268460.1	GI:3796064										
KEYWORDS	GSS.											
SOURCE	human.											
ORGANISM	Homo sapiens											
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;											
	Primates; Catarrhini; Hominoidea; Homo.											
	1 (bases 1 to 434)											
AUTHORS	Adams M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,											
	Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.											
	Use of human BAC End Sequences for Sequence-Ready Map Building											
TITLE	Unpublished (1998)											
JOURNAL												

LOCUS	AQ268460	434 bp	DNA	GSS	26-OCT-1998							
DEFINITION	RPC111-74G21.TJ RPC111 Homo sapiens genomic clone R-74G21, genomic survey sequence.											
ACCESSION	AQ268460											
NID	53796064											
VERSION	AQ268460.1	GI:3796064										
KEYWORDS	GSS.											
SOURCE	human.											
ORGANISM	Homo sapiens											
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;											
	Primates; Catarrhini; Hominoidea; Homo.											
	1 (bases 1 to 434)											
AUTHORS	Adams M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,											
	Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.											
	Use of human BAC End Sequences for Sequence-Ready Map Building											
TITLE	Unpublished (1998)											
JOURNAL												

Contact: Mark Adams

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel.: 301 838 0200
Fax: 301 838 0208

Email: mdadams@tigr.org
For clone availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from Research Genetics (inforesgen.com). BAC end search page: http://www.tigr.org/tdb/human/bac_end_search/bac_end_search.html.

```

FEATURES
  source      1.434
              Location/Qualifiers
Class: BAC ends.
Seq primer: SP6

```

```

/organism="Homo sapiens"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
/db_xref="GDB:7528196"
/db_xref="taxon:9606"
/clone="R-74G21"
/clone_lib="RPC11"
/sex="Male"
/cell_type="Lymphocytes"
147 a 74 c 89 g 124 t
BASE COUNT

```

ORIGIN	Query Match	21.1%	Score 24:	DB 40:	Length 434:
--------	-------------	-------	-----------	--------	-------------

Best Local Similarity 76.1%; Pred. No. 2.45e-02;
Matches 35; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 388 TGGATAAGATTAGATTTACGATGTGAATTAATAGAATATCTC 433
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 107 TGGGTAAGAAATTTCTTAAATCTTATGAATTAATTAGACTATTC 62

RESULT	11				
LOCUS		AU033574	435 bp	mRNA	EST 28-OCT-1998
DEFINITION		AU033574 Dictyostelium discoideum SL (H.Urushi-hara) Dictyostelium discoideum cDNA clone SUB183, mRNA sequence.			

ACCESSION AU033574
NID 93798998
VERSION AU033574.1 GI:3798998
KEYWORDS EST.

SOURCE
ORGANISM
DICTYOSTELIUM DISCOIDEUM.
DICTYOSTELIUM DISCOIDEUM
EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
1 (bases 1 to 435)
YOSHINO, R.; MORIO, T. and TANAKA, Y.
DEVELOPMENTAL cDNA IN DICTYOSTELIUM DISCOIDEUM
UNPUBLISHED (1997)
COMMENT
ON Jan 14, 1998 this sequence version replaced gi:1798828.

Contact: Hideo Urushihara
Institute of Biological Sciences
University of Tsukuba

ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigoidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 252)
AUTHORS Santos, T.M., Azevedo, V., Marotta, G.B., Santos, R.L., Fonseca, S.J.,
Ortega, J.M., Rabelo, E.M.L., Saber, M., Abdel-Hamid, H., Ridgers, I.L.,
Johnston, D.A., Fernandez, M., Rollinson, D., Franco, G.R. and
Pena, S.D.J.
TITLE Analysis of the Gene expression profile in Schistosoma mansoni
cercariae using EST approach
JOURNAL Unpublished (1998)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2043380.
Contact: Santos, T.M. and Pena, S.D.J.
Laboratorio de Genetica-Bioquimica, Departamento de Bioquimica e
Imunologia
Instituto de Ciencias Biologicas, Universidade Federal de Minas
Gerais
Avenida Antonio Carlos 6627, Belo Horizonte, MG, Brazil, 31270-010
Tel: (5531)4415611
Fax: (5531)4415409
Email: santostm@mona.icb.ufmg.br
Seq primer: M13 Forward
FEATURES
source
1..252
/organism="Schistosoma mansoni"
/strain="NMRI"
/note="Vector: pBluescript SK; Site_1: XhoI; Site_2:
ECORI; mRNA was extracted from cercariae and the library
was constructed and excised according to the
manufacturer's instructions (Uni-zap XR vector,
Stratagene)"
/db_xref="taxon:6183"
/clone="SM3650C"
/clone_lib="Schistosoma mansoni, cercariae"
/dev_stage="Larvae"
/lab_host="DH5alpha"
BASE COUNT 97 a 60 c 15 g 80 t
ORIGIN
Query Match 20.2%; Score 23; DB 21; Length 252;
Best Local Similarity 71.7%; Pred. No. 1.44e-01;
Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Db 161 CCACCAGGAATTTTTCATATAATTTTCTATATTCATCAT 213
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 53 CCACCGTTTATTTTGCATCTTTTCTGAAAGCTTTTAAATATCATCAT 1
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
RESULT 15
LOCUS AA334468 293 bp mRNA EST 21-APR-1997
DEFINITION EST38687 Embryo, 9 week Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA334468
NID 91986712
VERSION AA334468.1 GI:1986712
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 293)
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,
Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C.,
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S.,
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,
Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
Small, K.V., Spriggs, T.R., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,

He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,
Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H.,
Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M.,
Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,
Fraser, C.M. and Venter, J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
On Apr 14, 1993 this sequence version replaced gi:716849.
Other_ESTs: THC178342
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tldb/hg/hgi.html>)
Seq primer: M13 Reverse
FEATURES
Location/Qualifiers
1..293
/organism="Homo sapiens"
/note="Vector: pBluescript SK; Site_1: EcoRI; Site_2:
XhoI"
/db_xref="ATCC (inhost):136145"
/db_xref="taxon:9606"
/map="10"
/clone_lib="Embryo, 9 week"
/dev_stage="embryo, 9 wks"
BASE COUNT 79 a 51 c 46 g 116 t 1 others
ORIGIN
Query Match 20.2%; Score 23; DB 11; Length 293;
Best Local Similarity 76.7%; Pred. No. 1.44e-01;
Matches 33; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Db 234 ATAAATTTTAAAAATTTTAAAAAAGGTTTCGAAAAATGTAA 276
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 4 ATGATATTTTAAAAACCTTCAGAAAAAGAAATTCGAAAAAATAA 46
Search completed: Wed Sep 15 10:52:51 1999
Job time : 311 secs.

